(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12659 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/IB00/01496

(22) International Filing Date: 18 August 2000 (18.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,499 18 August 1999 (18.08.1999) US
60/156,503 28 September 1999 (28.09.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/156,503 (CIP)
Filed on 18 August 1999 (18.08.1999)

(71) Applicant *(for all designated States except US)*: GERMAN HUMAN GENOME PROJECT [DE/DE]; Fraunhofer Patentstelle, Leonrodstrasse 68, 80636 Munich (DE).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: WIEMANN, Stefan [DE/DE]; Grosse Lachstrasse 30a, 69207 Sandhausen (DE).

(74) Agent: CARPMAELS & RANSFORD; 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HUMAN DNA SEQUENCES

(57) Abstract: Novel human cDNA sequence of a clones, the encoded protein sequence of a clones, antibodies and variants thereof, are provided. The disclosed sequence of a clones find application in a number of ways, including use in profiling assays. In this regard, various assemblages of nucleic acids or proteins are provided that are useful in providing large arrays of human material for implementing large-scale screening strategies. The disclosed sequence of a clones may also be used in formulating medicaments, treating various disorders and in certain diagnostic applications.

WO 01/12659 A2

HUMAN DNA SEQUENCES

Background of the Invention

Current methods for testing pharmacological substances rely on a three-stage testing approach to drug development. First, candidate compounds are typically screened in some sort of *in vitro* system, like inhibition of cancer cell growth. Candidates are then tested in an animal model, as a first approximation of systemic effects, including efficacy and toxicity. Compounds that still show promise after these initial *in vivo* screens, finally are tested in humans. Again, human testing typically occurs in three phases: toxicity; preliminary efficacy; and efficacy. The entire process can take more than a decade and cost hundreds of millions of dollars. Aside from the monetary costs and protracted time scale, moreover, current testing regimes waste the lives of countless laboratory animals and needlessly endanger the lives of human subjects.

A need exists, therefore, for more sophisticated drug screening techniques that can be done rapidly *in vitro*. These screening techniques ideally will be reflective of systemic and/or organ-specific responses, so that they provide a reliable indicator of action in a human body. Current techniques, however, tend to utilize only a single or limited number of markers, thus answering only very simple questions that are of questionable medical import. For example, a typical *in vitro* assay may ask whether a lead compound binds a particular receptor, which has been implicated in a certain disorder. It is presumed that such binding is indicative of therapeutic usefulness, but it does not even purport to address systemic effects.

Not only are screening techniques for efficacy inadequate, the available toxicity screens likewise are inadequate. Toxicity, on a first level, is usually measured by animal testing. Aside from the complications related to *in vivo* versus *in vitro* testing, such screens are insufficient because of differences in metabolism, uptake, etc., relative to humans. Thus, improved methods would be not only be *in vitro*-based, they would also be more "human."

With the increasing miniaturization of screening assays and the growing availability of targets for pharmaceutical intervention, there is increasing interest in developing arrays containing large numbers of these targets that can be assayed simultaneously. If such an

array contains a large enough population of targets, it can be used to essentially mimic the systemic response. In other words, the array becomes an *in vitro* surrogate for the human body. The more refined the array, the more accurate the predictive capability. In theory, an array could be constructed that can detect all of the known human expression products simultaneously, thereby, providing a very reliable indicator of the human response to a given compound. These arrays offer advantages over the present *in vitro* screening systems in that they can assay large numbers of responses simultaneously. They are superior to animal testing because they are more "human" and, thus, more predictive of human responses.

In order to construct such arrays, however, the field is in need of further human targets. Advantageously, such targets will be provided with additional physiologically relevant information, such as whether the target is expressed in a particular tissue and whether it is related to a known functional class of targets. In this way, the artisan can focus as needed, for example, on tissue-specific effects or target class-specific effects, thereby providing information useful in evaluating efficacy and/or toxicity.

In addition to a need for pharmacological screening targets, there is a need for further pharmacological substances. These substances can be used in the formulation of medicinal compositions and in treating a wide variety of disorders.

The present invention responds to the aforementioned and other needs in the field by providing a population of novel targets useful, *inter alia*, in the profiling and medicinal contexts described above.

Summary of the Invention

It is an object of the invention, therefore, to provide a set of human cDNA clones. Further to this object, the invention provides sequences of human cDNA clones that were isolated from libraries generated from different human tissues.

It is another object of the invention to provide assemblages of targets useful in profiling matrices for screening pharmacological test compounds. According to this object, assemblages comprising different populations of human nucleic acids, proteins and antibodies are provided. In different embodiments, cDNA library-specific assemblages and target-family-specific targets are provided.

It is a further object of the invention to provide a database of human nucleotide and protein sequences. Further to this object, novel human nucleotide and protein sequences are provided in electronic form. In one embodiment, one or more of these sequences is provided in a searchable database.

It is still another object of the invention to provide biologically active target molecules useful in treating or detecting human disorders. Further to this object, the invention provides nucleic acid and protein molecules that have the capacity to affect disease etiology or symptoms or correlate with known disease states. Also further to this object, a database is provided which comprises the disclosed molecules in electronic form.

It is still a further object of the invention to provide polypeptides encoded by the human cDNA clones disclosed herein. Further to this object, the invention provides antibodies and fragments thereof that are capable of binding to a specific portion of these polypeptides.

It is yet another object of the invention to provide pharmaceutical compositions which comprise an effective amount of a pharmaceutical agent, wherein the pharmaceutical agent is selected from the group consisting of one or more polypeptides contemplated by the invention, variants or functional derivatives thereof, and antibodies thereto; and a physiologically acceptable carrier or excipient.

It is still another object of the invention to provide expression vectors comprising one or more human cDNA clones disclosed herein or fragments thereof; and optionally a promoter operably linked to the cDNA clone or fragment thereof. Further to this object, the invention provides methodology for recombinantly producing a desired peptide, comprising expressing in a host cell a peptide encoded by a human cDNA clone disclosed herein.

Detailed Description

The invention results from a need in the art for new human nucleic acids and proteins. This need arises in several contexts. First, there is a need to identify targets for therapeutic intervention. Second, there is a need to identify molecules that may be adversely affected in a therapeutic context, thereby resulting in toxicity. Knowledge of these molecules will aid in

the design of new medicaments with enhanced efficacy and decreased toxicity. Finally, the need encompasses human nucleic acids and proteins that have medicinal applicability in their own right.

In view of these needs, the present inventors set out to isolate and sequence human cDNAs from tissue-specific libraries. In this way, they represent subsets of molecules likely to be targets for therapeutic intervention or for avoiding toxicity. In addition, the inventors divided the molecules into various sub-categories, based on suspected functionality, structural similarity etc, which are of interest from a pharmacological perspective. These molecules are disclosed in provisional application serial nos. 60/149,499 and 60/156,503, filed August 18, 1999, and September 28, 1999, respectively, both of which are hereby incorporated by reference in their entirety.

GENERAL DESCRIPTION OF THE INVENTIVE MOLECULES

The present invention provides novel polynucleotide molecules that, in some instances, have similarities with known molecules. The inventive DNAs were cloned from five different human cDNA libraries. In addition to these DNA molecules, the invention provides their protein translations and antibodies derived from them. The inventive DNA and protein sequences are show individually, below. The inventive nucleic acids also include the complements of these DNA sequences, as well as their RNA counterparts. Methods of producing the molecules also are provided. Further, the invention provides methods for detecting all or part of the molecules and of detecting polynucleotides encoding all or part of the molecules.

The inventive molecules derive from five cDNA libraries: human fetal brain; human fetal kidney; human mammary carcinoma; human testis; and human uterus. For convenience, each sequence bears a designation that indicates from which library it is derived. In particular, these designations are: “hfpbr” for human fetal brain; “hfkd” for human fetal kidney; “hmcf” for human mammary carcinoma; “htes” for human testis; and “hute” for human uterus. The individual libraries were constructed and screened as described below in the examples.

The protein and DNA molecules of the invention are variously described herein as "target" molecules or "inventive" molecules. The sequences and other information pertinent to the nucleic acid and protein molecules of the invention are shown, below.

Interpreting the data disclosed with the Table and cDNA sequences, below:

The table and data below provide the coding sequences of the inventive cDNAs as well as the protein sequences and other useful information, as set out below.

Grouping

The clones were assigned to the following fourteen functional and/or tissue-derived groups:

1. Cell Cycle
2. Cell Structure and Motility
3. Differentiation/Development
4. Intracellular Transport and Trafficking
5. Metabolism
6. Nucleic Acid Management
7. Signal Transduction
8. Transmembrane Protein
9. Transcription Factors
10. Brain derived
11. Kidney derived
12. Mammary Carcinoma derived
13. Testes derived
14. Uterus derived

Description of Clone Files

The individual clone files are structured in the same pattern. The Sections are separated by paragraphs.

1. Clone Name

The clone names are deciphered with reference to the following example: DKFZphfkd2_24e23, wherein the code represents:

- producer of library ("DKFZ") (for convenience, this reference may be eliminated)
- a "p" for "plasmid cDNA library" (for convenience, this reference may be eliminated)
- library name (e.g. hfbr = human fetal brain; hfkd = human fetal kidney; hmcf = human mammary carcinoma; htes = human testes; hute = human uterus)
- an underscore ("_") to separate library information from plate information
- plate number (e.g. "16")
- plate coordinates (letter first; e.g. "f14")

2. Group

3. Introduction

short review of the similarities, function of the protein and possible applications

4. Short Information

specifications about the cDNA (who sequenced, completeness of the cDNA, similarity, who sequenced, chromosomal localisation, length of cDNA, localisation of poly A tail and polyadenylation signal)

5. cDNA-Sequence

6. BLASTn Results

search results of blasting the cDNA sequence against all public databases

7. Medline Entries

information about genes/proteins similar to the novel cDNA (if available)

8. Putative Encoded Protein Information

specifications about the encoded protein (ORF: length and localisation of the reading frame)

9. Protein Sequence

10. BLASTp Results

search results of blasting the protein sequence against all public databases

11. Pedant Information

output of fully automated annotation: summarises peptide information, homologies, patterns as follows:

[Length]

- length of the protein = number of amino acid residues

[MW]

- molecular weight of the protein

[pI]

- isoelectric point

[HOMOL]

- shows protein with closest similarity to the cDNA-encoded protein

[FUNCAT]

- functional information according to a catalogue developed by Munich Information center for Protein Sequences (MIPS)

[BLOCKS]

- Blocks are multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins. The blocks for the Blocks Database are made automatically by looking for the most highly conserved regions in groups of proteins documented in the Prosite Database. The Prosite pattern for a protein group is not used in any way to make the Blocks Database and the pattern may or may not be contained in one of the blocks representing a group. These blocks are then calibrated against the SWISS-PROT database to obtain a measure of the chance distribution of matches. It is these calibrated blocks that make up the Blocks Database. The WWW versions of the Prosite and SWISS-PROT Databases that are used on this server are located at the ExPASy World Wide Web (WWW) Molecular Biology Server of the Geneva University Hospital and the University of Geneva. World Wide Web URL http://blocks.fhcrc.org/blocks/about_blocks.html/ is the entry point to the database.

- here Blocks segments found in the analysed protein sequences are displayed

[SCOP]

Nearly all proteins have structural similarities with other proteins and, in some of these cases, share a common evolutionary origin. The scop database provides a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known, including all entries in Brookhaven National Laboratory's Protein Data Bank (PDB). It is available as a set of tightly linked hypertext documents which make the large database comprehensible and accessible. In addition, the hypertext pages offer a panoply of representations of proteins, including links to PDB entries, sequences, references, images and interactive display systems. World Wide Web URL http://scop.mrc-lmb.cam.ac.uk/scop/ is the

entry point to the database. Existing automatic sequence and structure comparison tools cannot identify all structural and evolutionary relationships between proteins. The scop classification of proteins has been constructed manually by visual inspection and comparison of structures, but with the assistance of tools to make the task manageable and help provide generality. Proteins are classified to reflect both structural and evolutionary relatedness. Many levels exist in the hierarchy, but the principal levels are family, superfamily and fold. The exact position of boundaries between these levels are to some degree subjective. Scop evolutionary classification is generally conservative: where any doubt about relatedness exists, we made new divisions at the family and superfamily levels.

- - here SCOPE segments found in the analysed protein sequences are displayed

[EC]

ENZYME is a repository of information relative to the nomenclature of enzymes. It is primarily based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) and it describes each type of characterized enzyme for which an EC (Enzyme Commission) number has been provided. World Wide Web URL http://www.expasy.ch/enzyme/ is the entry point to the database.

- here EC-number and name of enzymes with similarity to the analysed protein sequences are displayed

[PIRKW]

- functional information according to the Protein Information Resource (PIR) database catalogue developed by Munich Information Center for Protein Sequences (MIPS), the National Biomedical Research Foundation (NBRF) and the International Protein Information Database in Japan (JIPID).

[SUPFAM]

- information according to the Protein Information Resource (PIR) database catalogue of protein superfamilies developed by Munich Information Center for Protein Sequences (MIPS), the National Biomedical Research Foundation (NBRF) and the International Protein Information Database in Japan (JIPID).

[PROSITE]

please refer to 12. PROSITE Motifs

[PFAM]

please refer to 13. PFAM Motifs

[KW]

- overall 2dimensional folding information

- 3D indicates that the proteins is similar to a protein of which a 3 dimensional structure is known

- overall structural information

[]

The last PEDANT-block depicts information about the folding structure of the protein generated by PREDATOR. PREDATOR is a secondary structure prediction program. It takes as input a single protein sequence to be predicted and can optimally use a set of unaligned sequences as additional information to predict the query sequence. The mean prediction accuracy of PREDATOR is 68% for a single sequence and 75% for a set of related sequences. PREDATOR does not use multiple sequence alignment. Instead, it relies on careful pairwise local alignments of the sequences in the set with the query sequence to be predicted.

World Wide Web URL http://www.embl-heidelberg.de/argos/predator/predator_info.html is the entry point to the database.

- H = helix, E = extended or sheet, _ = coil, T = transmembrane, B = beta

- x indicates a low-complexity region with repeat-like structure which is omitted in all BLAST searches

12. PROSITE Motifs

PROSITE is a database of protein families and domains. It consists of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family (if any) a new sequence belongs. World Wide Web URL http://www.expasy.ch/prosite/ is the entry point to the database. A description of the prosite consensus patterns is also provided, below.

13. PFAM Motifs

PFAM (protein families) is a large collection of multiple sequence alignments and hidden

Markov models covering many common protein domains. World Wide Web URL http://www.sanger.ac.uk/Pfam/ is the entry point to the database.

Deposit of Clones

Clones were deposited as a pool with the American Type Culture Collection under accession number __________, from which each clone comprising a particular polynucleotide is obtainable. Each clone has been transfected into separate bacterial cells (E. coli) in this composite deposit.

The clones may also be obtained from the Resource Center of the German Human Genome Project (Heubner Weg 6, 14059 Berlin, GERMANY). The Resource Center library numbers are slightly different that those presented here, but may be readily obtained by the following key or with the assistance of Resource Center personnel.

The library name becomes a number: brain (hfbr2) becomes 564; kidney (hfkd2) becomes 566; mammary carcinoma (hmcf1) becomes 727; testis (htes3) becomes 434;and uterus (hute1) becomes 586. Next, the plate number is converted to two digits (e.g., "2" becomes "02") and is moved behind the plate coordinate, and the underscore is dropped. The following examples are helpful:

Listed Number	Resource Center Number
DKFZphfbr2_16f21	DKFZp564F2116
DKFZphfkd2_1j9	DKFZp566J091
DKFZphmcf1_1c23	DKFZp727C231
DKFZphtes3_14g5	DKFZp434G0514
DKFZphute1_17k7	DKFZp586K0717

The libraries were constructed using two commercially available vectors. The brain (hfbr2 designations) and kidney (hfkd2 designations) libraries utilize pAMP 1 from Life Technologies and are maintained in XL-2Blue (Strategene); the uterus (hute1), testes (htes3) and mammary carcinoma (hmcf1) libraries are constructed in pSPORT1, also from Life Technologies, and are maintained in DH10B (LifeTechnologies). In addition to the following techniques, consultation with the commercial literature available on these clones will make evident all of the housekeeping techniques needed to propagate and isolate the individual constructs. All inserts may be excised with a NotI/SalI digestion. Alternatively, universal primers, flanking the cloning region, may be used to amplify the inserts using PCR methods.

Bacterial cells containing a particular clone can be obtained from the composite deposit as follows:

An oligonucleotide probe or probes should be designed to the sequence that is known for that particular clone. This sequence can be derived from the sequences provided herein, or from a combination of those sequences. Methods of probe design are presented below.

Oligonucleotide probes may be labeled with γ-^{32}P ATP (specific activity 6000 Ci/mmole) and T4 polynucleotide kinase using commonly employed techniques for labeling oligonucleotides. Other, non-radioactive labeling techniques can also be used. Unincorporated label typically is removed by gel filtration chromatography or other established methods. The amount of radioactivity incorporated into the probe can be quantified by measurement in a scintillation counter. Preferably, specific activity of the resulting probe generally should be approximately $4X10^6$ dmp/pmole.

The bacterial culture containing the pool of full-length clones should preferably be thawed and 100 µl of the stock used to inoculate a sterile culture flask containing 25 ml of sterile L-broth containing ampicillin at 50 - 100 µg/ml (for XL-2Blue strains 25 µg/ml tetracycline should also be used). The culture should preferably be grown to saturation at 37°C., and the saturated culture should preferably be diluted in fresh L-broth. Aliquots of these dilutions should preferably be plated to determine the dilution and volume which will yield approximately 5000 distinct and well-separated colonies on solid bacteriological media containing L-broth containing ampicillin at 100 µg/ml (for XL-2Blue strains 25 µg/ml tetracycline should also be used)and agar at 1.5% in a 150 mm petri dish when grown overnight at 37°C. Other known methods of obtaining distinct, well-separated colonies can also be employed.

Standard colony hybridization procedures should then be used to transfer the colonies to nitrocellulose filters and lyse, denature and bake them. The filter is then preferably incubated at 65°C. for 1 hour with gentle agitation in 6 x SSC (20 x stock is 175.3 g NaCl/liter, 88.2 g Na citrate/liter, adjusted to pH 7.0 with NaOH) containing 0.5% SDS, 100 µg/ml of yeast RNA, and 10 mM EDTA (approximately 10 mL per 150 mm filter). Preferably, the probe is then added to the hybridization mix at a concentration greater than or equal to $1X10^6$ dpm/mL. The filter is then preferably incubated at 65°C. with gentle agitation overnight. The filter is then preferably washed in 500 mL of 2 x SSC/0.5% SDS at room temperature without agitation, preferably followed by 500 mL of 2 x SSC/0.1% SDS at room

temperature with gentle shaking for 15 minutes. A third wash with 0.1 x SSC/0.5% SDS at 65°C. for 30 minutes to 1 hour is optional. The filter is then preferably dried and subjected to autoradiography for sufficient time to visualize the positives on the X-ray film. Other known hybridization methods can also be employed.

The positive colonies are picked, grown in culture, and plasmid DNA isolated using standard procedures. The clones can then be verified by restriction analysis, hybridization analysis, or DNA sequencing.

Alternatively, clones may be grown as described above, and PCR used to isolate the insert DNAs. Methods of PCR are described below and are otherwise well known .

ERROR SCREENING

The DNA sequences found herein derive from individual clones, which are publicly available, as noted above. Thus, the skilled artisan will recognize that any specific sequence disclosed herein readily can be screened for errors by resequencing a particular fragment, in both directions (*i.e.*, by sequencing both strands). Alternatively, error screening can be performed by amplifying and/or cloning any of the inventive DNAs, using for example RT-PCR, and sequencing the resulting amplified product. In the event that there is a sequencing error, reference should be made to the deposited clone as the correct sequence.

USES AND BIOLOGICAL ACTIVITIES OF THE INVENTIVE MOLECULES

The inventive molecules and their derivatives are susceptible to a wide variety of uses, based on functional and/or structural properties. The skilled worker will appreciate, based on the biological activities detailed below, and discussed with regard to the individual sequences disclosed below, that the inventive molecules will find usefulness in numerous therapeutic and diagnostic applications.

The DNA molecules, especially the potassium salts thereof, can be used as fertilizer supplements due to their high nitrogen and phosphorus contents. Since the DNAs are of defined length, they are also useful in gel electrophoresis as molecular weight markers. Due to their similarity with known molecules, certain of the DNA molecules and their variants and derivatives may be used in any number of different diagnostic procedures and therapeutic applications. They may also be used to make the encoded proteins.

The proteins themselves have many possible uses. They may be used as a nutritional supplement for humans, animals and even for laboratory use as, for example, medium for bacterial cultures. Moreover, since the proteins are of defined, known sizes, they may be used as molecular weight markers for gel electrophoresis and gel filtration. Because they are of defined sequences, they also have use in microsequencing and protein fingerprinting applications.

Expression Profiling Applications

Given their known tissue expression and functional associations, assemblages of the inventive proteins (or corresponding antibodies) and nucleic acids are particularly suited to expression profiling applications. Expression profiling generally entails constructing an array of indicators that signal the presence of a particular RNA or protein expression product. Such arrays can be used to evaluate, for example, pharmacological effectiveness and toxicity. In particular, expression profiles from such arrays can be generated from cells treated with known compounds, having known properties, and these profiles can be compared to profiles of unknowns to evaluate similarities and differences, which can be correlated with efficacy or toxicity.

Additional uses of profiling include diagnosis, tracking development, and ascertaining signaling and metabolic pathways. For examples of references describing profiling and its uses, see Farr *et al.*, U.S. Patent 5,811,231 (1998); Seilhamer *et al.*, U.S. Patent 5,840,484 (1998); Rine *et al.*, U.S. Patent No. 5,777,888 (1998); WO 97/27317; WO 99/05323; WO 99/09218; and WO 99/14369. For a device for implementing such techniques, see Lipshutz *et al.*, U.S. Patent No. 5,856,174 (1999) and Anderson et al., U.S. Patent No. 5,922,591 (1999).

In one embodiment, a subset of the inventive DNAs will be arrayed on a substrate, like a gene chip, a filter or a 96-well plate. Test samples containing cells are maintained in the presence of a label capable of incorporation into nascent mRNA. Samples are treated with test and control compounds, which will induce mRNA expression in the sample, resulting in incorporation of label. Whole mRNA is isolated and applied to the array such that it hybridizes with the DNAs contained therein. After washing, the amount of hybridization is quantified and a profile is generated. These steps are repeated with various control and test compounds, thereby generating a library of profiles, which can be used to ascertain the relationships relevant to pharmacological efficacy or toxicity.

The matrices used in such profiling, however, need not be limited to those utilizing DNAs. Rather, other nucleic acids, like RNAs and protein nucleic acids (PNAs), as well as the inventive proteins and antibodies corresponding to the inventive proteins may also be employed. Hence, for example, antibodies could form the array and the samples could be treated in order to label nascent proteins. Whole proteins then would be isolated and applied to the antibody matrix. Developing the resulting signal would result in a protein expression profile, which is useful in essentially the same manner as the nucleic acid profile. A protein matrix could be used, for example, in evaluating antibody responses to pharmaceutical agents in order to eliminate possible cross-reactivity.

Moreover, where nucleic acids are used in the matrix, it is often beneficial to use variants (as defined below) of the molecules described herein. This can be used to account for genetic variations that are of little or no consequence to the function of the resultant gene product. Hence, they can account for wobble or conservative amino acid variations that do not perturb function, like variations in some of the protein motifs elucidated below. Thus, each position in the matrix can employ multiple nucleic acid probes that account for a series of variants.

Expression profiling may also be done, in another embodiment, using two-dimensional protein gels in which the inventive proteins are detected. The resultant profiles can be used in the same way as described.

Matrices useful for profiling may be constructed based on different criteria. Of course, the more relevant profiles will take into account expression of most human genes, preferably all of them. In certain situations, however, it is advantageous to look at a smaller subset. For example, if one were concerned about fetal neural toxicity, a fetal brain-specific matrix might be chosen. On the other hand, if one were interested in targeting mammary carcinoma tissue, a corresponding matrix could be used. Thus, matrices may be constructed using all of the sequences available from a tissue-specific library.

* * *

The following discussion relates to some of the various functional and structural groupings that would be of interest to the artisan wishing to construct profiling matrices. Of course, the artisan will also recognized that these functional descriptions may find additional applicability in the therapeutic and diagnostic applications discussed below.

<u>Cell Cycle</u>

A proliferating cell must coordinate replication and chromosomal separation to ensure that the genome is replicated completely, and that a single copy is correctly inherited by each daughter cell. The cell cycle is the coordinated series of events that achieves these aims. Many of the key events are initiated by a family of conserved Seiren/threonine protein kinases, the cyclin-dependent kinases (CDKs), that are activated by the cyclin family of proteins (cyclins A-H). In turn, the cyclin-CDK complexes are modulated by other protein kinases or phosphatases, and by binding specific inhibitor proteins. The enormous variety of ways in which CDK activity can be regulated allows the cell to respond to internal signals generated by preceding events in the cell cycle and to external growth signals.

The somatic cell cycle is divided into four phases: DNA replication (S phase) and chromosome separation (M phase) are separated by gap phases (G1 and G2). At specific control points the decision to begin the next stage (DNA synthesis or mitosis) is carefully regulated.

Cdc2, the primary kinase, is especially required for the G1-S transition and S phase. Cdc4 and Cdc6 are involved at the restriction point, where the cell can decide to proliferate or arrest (G1<->G0) and Cdc7 is a CDK activating kinase (CAK) as well as a subunit of TFIIH.

The Cyclin-CDK complexes are regulated in various ways. One is through phosphorylation by CDK activating kinases (CAK), like the Y15 kinase (Wee1) and dephosphorylation by CDK associated phosphatases (CAP), like Cdc25A a member of the Cdc25 family (Cdc25A, B and C).

An other way of regulation occurs through two classes of CDK inhibitors (CKI), the INK4 proteins p15, p16, p18, and p19, who negatively regulates the cyclin D CDK complexes and second the p21 family with p21, p27, and p57.

The cell cycle is also regulated through ubiquitin-mediated proteolysis involving the destruction of both cyclins and CDK inhibitors by the 26S proteasome, that requires an ubiquitin conjugating enzyme (UBC) and an ubiquitin ligase. The instability is conferred by PEST regions (cyclin D and E) or a ten amino acid region in the amino terminus (degradation box) in the A- and B-type cyclins.

All these modifications play an important role for the cellular localization, because only the nuclear CDK-cyclin complexes are functional for cell cycle. During G1 phase of the cell cycle, cyclines A, E and D are synthesized and bind to their cyclin-dependent kinase (CDK) partners. CDK complexes containing cyclins A, E and D1 are then imported into and concentrated within nuclei. Cdk6- cyclin D3 has been localized to both cytoplasmic and nuclear compartments, although only the nuclear complex is active. As cells enter S phase, cyclin A and cyclin E complexes remain within the nucleus, whereas cyclin D1 relocalizes to the cytoplasm for proteolysis at the onset of S phase. Like Cdk2-cyclin A, Cdc2-cyclin A is nuclear and remains so until it is degraded during mitosis. By contrast, as a result of ongoing nuclear import and more rapid re-export, cyclin B1, which binds to Cdc2 upon synthesis during S phase, is predominantly cytoplasmic. Cdc2-cyclin B2 is also cytoplasmic, although this might occur through anchoring of the complex to some cytoplasmic constituent. At prophase, phosphorylation of cyclin B1 promotes accumulation of Cdc2-cyclin B1 in the nucleus, whereas cyclin B2 remains in the cytoplasm until nuclear envelope breakdown.

Two crucial regulators of Cdc2-cyclin B-Wee1 and Cdc25C exist and are responsible for the G2 to M control point. Wee1 is a nuclear protein throughout the cell cycle, whereas Cdc25C binds to 14-3-3 proteins during interphase and remains predominantly cytoplasmic. In some systems Cdc25C, like cyclin B1, rushes precipitously into the nucleus just before entry into mitosis.

The 110-kDa retinoblastoma (tumor suppressor) protein (RB), a pRB-family member is an important regulator of cell-cycle progression and differentiation. Like the E2F family (E2F1-5) or DP family (DP1-3) of transcription activators, RB suppresses inappropriate proliferation by arresting cells in G1 by repressing the transcription of genes required for the transition into S phase. Before the cell proceeds into S phase, RB becomes phosphorylated at multiple sites by the cyclin dependent protein kinases (CDKs) and loses its transcriptional repressing activity. Phosphorylation of RB during late G1 phase results in the dissociation of the E2F-RB repressor complex which allows S-phase specific genes to be transcribed. Cyclin E is the evolutionary conserved target for E2F and interacts together with CDC2 in late G1.

For a proliferating cell it is vital that only undamaged DNA is replicated because if DNA damage is substantial, its replication can lead to chromosome loss or rearrangement.

Thus, we find a G1<->S checkpoint in late G1 that requires tumor suppressor p53. A p53-dependent G1 arrest is effected by the cyclin dependent kinase inhibitor p21 through higher expression levels that inhibits almost all cyclin CDK complexes.

The kinase responsible for phosphorylating the unidentified kinetochore component in metaphase may be a member of the MAP kinase family and appears to be the proto oncogene c-MOS, a cytostatic factor (CSF) in meiosis.

Several categories of proteins are coded for by clones of the invention within the overall group of "Cell cycle"and include, among others, the following:

Tumor suppressors (e.g. N33): Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. The N33 gene has been reported by OMIN OMIN (Online Mendelian Inheritance in Man at http://www.ncbi.nlm.nih.gov/htbin-post/Omin) to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) prostate cancer suppression (OMIN *601385). Clones in this category include: fbr2_2k14.

C-TAK1 Cdc25c associated protein kinase: Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five C associated protein kinase) phosphorylates Cdc25C on serine 216 in vitro. Alterations in the gene coding for the above protein kinase has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with Pancreatic cancer (OMIN *60278). Clones in this category include: tes3_7j3.

Cell structure and motility

One of the major differences between prokaryotes and eukaryotes is the ability of the eukaryotic cell to adopt very different shapes dependent on its function during the differentiation process. Animal cells vary from being round to extended cylindric forms like motorneurons or muscle cells. In humans, more than 100 different cell types can be distinguished, each having a characteristic shape. The form of a cell often is closely related to

its capacity to move. Some completely differentiated cells like fibroblasts can still change their form actively, thereby migrating. Other cell types serve as motor elements - "macroscopically" like muscle cells or "microscopically" like ciliated epithelia. Such tasks are fulfilled by a big class of proteins; on the one hand responsible for maintenance of cell structure and contacting neighbor cells or the intercellular matrix and on the other hand for cell motility. These topics cannot be regarded separately: The motility apparatus e.g. must be fixed in the cytoskeleton. Three different types of filaments can be distinguished: Actin filaments, tubulin filaments and intermediate filaments, each present in almost all types of cells.

Actin filaments (F-actin) are built up of monomers (G-Actin). In muscle cells, actin, myosin, for both of which several paralogous genes are known, as well as many more proteins are constituents of the contractile apparatus.

The "thin" and "thick filaments" in a muscle cell consist mainly of actin and myosin, respectively.

Several different proteins are responsible for the anchoring of the actin filaments in the Z-disks (e.g. alpha-actinin and desmin) or at the end of the myofibers in the cell membrane.

Troponin I, -C, -T and Tropomyosin - associated with actin - confer the Ca++-dependent triggering of contraction.

Length of the sarcomere is controlled by the giant protein titin.

In smooth muscle, there is no troponin. Contraction activity is controlled by phosphorylation / dephosphorylation of myosin by a specialized kinase instead. Contractile fibers are not organized in sarcomeres.

Apart from contributing to muscle contraction, the actomyosin system is responsible for many other motions at cellular level, e.g. the amoeboid movement of pseudopodia or the fission of cells at the end of mitosis by a contractile ring.

Besides this, actin fibers fulfill structural tasks like maintenance of the shape of stereocilia or microvilli. Here, actin filaments are connected by proteins like fimbrin. But not

only specialized structures like the mentioned ones contain actin fibers. There is a network covering the complete cell volume with F-actin as a major constituent. Whereas the actin filaments in the structures mentioned above are relatively stable, this F-actin is highly dynamic. Management of the network structure and turnover is achieved by connecting proteins like alpha-actinin, fimbrin or fill-in; turnover is regulated by gelsolin, villin, and different capping- and fragmentation-proteins.

Microtubules are built up of alpha-beta tubulin heterodimers. Turnover of filaments is achieved by building-in and releasing of monomers with different time constant rates at both ends. The resulting cycle is called "treadmilling". Thirteen strings of tubulin duplets build up one subfiber, whereas one fiber contains two or three of those. A complete axoneme consists of 9 radial and 2 central fibers. This "9+2" - structure is the basis both of flagella, their basal bodies and centrioles. In flagella, several additional structures like radial elements exist. Nexin connects the fibers and dyneine is the motor ATPase which shifts the fibers relative to each other. Several genetic diseases like the Cartageneric syndrome are caused by deficiencies of distinct proteins in cilia.

Besides this, microtubules are abundant in all types of cells. They are part of a delivery system for organelles, e.g. in the golgi apparatus. A further very important system based on microtubules is the mitotic spindle, it is organized by the centrosomes. Besides many other components, the major part of a centrosome are two centrioles which are built up of nine microtubule-triplets. Most remarkably, new centrioles are not synthesized de novo but generated by duplication of old ones.

Cytoplasmic microtubules are associated with many different proteins. Two major classes are known: The MAPs ("microtubule-associated proteins", with molecular masses between 200 and 300 kD) and the much smaller tau-Proteins with a MW between 60 and 70 kD. These proteins regulate the treadmill-process and the interaction with other structures in the cell.

Besides actin and myosin the so-called intermediate filaments constitute a third class of filaments. In contrast to the former two groups, they do not participate in motility, nor are they dynamic structures subject to a vivid turnover. The most important ones are

neurofilaments (in neurons), keratin filaments (mainly in epithelial cells), and vimentin filaments (in many sorts different cell types).

The biological function of both the cytoskeleton as well as contractile apparatus of a cell does not end at the cell membrane. Cells must be embedded in the extracellular matrix, all cells of a muscle must act as one single mechanical unit and epithelia must resist macroscopic mechanical forces. Hence, cell adhesion and the extracellular matrix are closely connected to the cytoskeleton. Vincullin is one of the proteins which serve as an anchor for intracellular fibers (actin). Different types of desmosomes and tight junctions connect neighbor cells with intercellular fibers. On the inside, cytoplasmic plaques connect them to the cytoskeleton. These structures, on the one hand, serve as mechanical elements whereas gap junctions, on the other hand, connect cells metabolically.

The extracellular matrix consists of a network of proteins, glycoproteins and polysaccharides. Different proteins are present in relation to different mechanical demands:. Elastin is found in tissues with high elasticity (lungs, heart) whereas collagen, a more hard-wearing protein, is found in tendons and ligaments. Fibronectin is an extracellular protein highly important for cell adhesion.

Reference: Murray J *et al* (1992): Cell Motil Cytoskeleton 22: 211-223.

Within the overall group of Cell Structure and Motility several categories of proteins are coded for by clones of the invention:

<u>Collagen alpha chain proteins</u>: Proteins with the typical (xxG)n repeat of collagen proteins and Pfam von Willebrand factor type A domain(s) suggest they are collagen alpha chains. These proteins can find application in modulation of connective tissue, bone and cartilage development and maintainance. OMIN reports collagen alpha chains have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Osteogenesis imperfecta, type I (OMIN #166200); 2) Osteogenesis imperfecta congenita (OMIN #166210); 3) Alport Syndrome, X-linked (OMIN #301050); 4) Thrombastenia of Glanzmann and Naegeli (OMIN *273800); 5) Ehlers-Danlos Syndrome, Type VII (OMIN #130060); 6) Marfan Syndrome (OMIN #154700); 7) Alport Syndrome, Autosomal Recessive (OMIN #203780); 8) Alpha-2-Deficient Collagen Disease (OMIN 203760); 9) Goodpasture Syndrome (Omin 233450); 10) Osteogenesis Imperfecta,

progressively deforming, with normal sclerae (OMIN #259420); 11)) Ehlers-Danlos Syndrome, Type VII Autosomal Recessive (OMIN *225410); and 12)) Osteogenesis imperfecta, Type IV (OMIN #166220). OMIN reports that von Willebrand factor type A domains have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases:: 1) Hemophilia A (OMIN *306700); 2) Von Willebrand Disease (OMIN *193400); 3) Giant Platelet Syndrome (OMIN *231200); 4) Thrombastenia of Glanzmann and Naegeli (OMIN *273800); 5) Congenital Thrombotic Diseasae due to protein C deficiency (OMIN #176860); 6) Polycystic Kidney Disease 1 (OMIN *601313); 7) Nephrogenic Diabetes Insipidus (OMIN *304800); 8) Factor V Deficiency (OMIN *227400); and 9) Dentatorubral-Pallidoluysian Atrophy (Omin *125370). Clones in this category include: fbr2_2b5.

Radial spokehead protein: Radial spokehead proteins, e.g., Chlamydomonas reinhardtii radial spokehead protein of flagella or axoneme and the Strongylocentrotus purpuratus sea urchin spermatozoa protein p63, and human proteins with similarity thereto are important for the maintenance of a planar form of sperm flagellar beating. The human protein(s) can find application in modulating the structure of the human spermatozoa radial spoke head and modulation of sperm motility in men (e.g., in sterility). Clones in this category include: tes3_15i5.

Ankyrins: Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus these proteins are involved in coupling of cyto skeleton and cell membrane. OMIN reports that Ankyrins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Heriditary Spherocytosis (OMIN *182900); 2) Hemolytic Poikilocytic Anemia due to reduced ankyrin binding sites (OMIN 141700); 3) Atypical Elliptocytosis (OMIN 225450); 4) Autosomal recessive spherocystosis (OMIN #270970); 5) Werner Syndrome (OMIN *277700); and 6) Rhesus-unlinked type Elliptocytosis (OMIN #130600). Clones in this category include: tes3_18l7.

FGD1-related F-actin binding protein (Farbin/FGD1): FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin-binding protein. The gene locus fgd1 seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. (OMIN 305400). Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as

described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an esin yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. Clones in this category include: tes3_72k15.

Paramyosins: Paramyosin is a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as Schistosoma mansoni. Clones in this category include: tes3_7b22.

Tuftelin: Tuftelin/enamelin are matrix proteins of the teeth. As other proteins involved in calcification, these proteins are also expressed in the uterus matrix. The new protein can find application in modulation of tissue-calcification, especially the uterus. As reported by OMIN, tuftelin has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with amelogenesis imperfecta (OMIN *600087). Clones in this category include: ute1_19g22.

Cell Adhesion Regulator (CAR1): CAR1 is involved in the regulation of cell-cell adhesion. OMIN reports the association (as potentially diagnostic, therapeutic, causative, and/or related, etc...) of CAR1 with tumor suppression by the reduction of tumor invasion (OMIN *116935). Clones in this category include: ute1_24j6.

Differentiation/Development

Almost every multicellular organism originates from meiotic cell divisions and the recombination of a paternal and a maternal set of chromosomes. After fertilization of the egg, all cells of a body originate from this one cell. Thus the cells of the developing body are initially genetically alike. But phenotypically they become very different. They are specialized to a certain cell type and arranged in an organized pattern to a certain type of tissue and the whole structure has the well-defined shape of an organ. All these features are determined by the DNA sequence of the genome, which is reproduced in every cell. Each cell acts on the genetic instructions given to a certain time and at a certain place of development and plays its individual part in the multicellular organism. Cell differentiation may be divided into three general steps: cell cycle exit, apoptosis protection and tissue specific gene

expression. These processes are coordinated to provide the final and unique tissue characteristics.

An animal cell that has achieved a certain level of development is said to be determined. This differentiation of a cell may be irreversible and in that case the cell may be renewed only by simple duplication. Other cells are renewed by means of stem cells which are immortal (e.g. stem cells of the bone marrow, epidermal stem cells). The genetic control of development is extensively studied in non-vertebrates and vertebrates. The classical animal model is the fruit fly Drosophilia and the modern model is the transgenic mouse. Animal transgenesis has proven to be useful for physiological as well as physiopathological studies. Besides the approach based on the random integration of a DNA construct in the mouse genome, gene targeting can be achieved using totipotent embryonic stem cells for targeted transgenesis. Transgenic mice are than derived from the embryonic stem cells. This allows the introduction of null mutations in the genome (so-called knock-out) or the control of the transgene expression by the endogeneous regulatory sequence of the gene of interest (so-called knock-in). Mice can be created that express wild-type genes, mutant genes, marker genes or cell lethal genes in a tissue specific manner. These animal models allow to follow changes in tissue and organ development and lead to a better understanding of the cellular function of many genes or to the generation of animal models for human diseases. Fundamental problems in immunology, onset and development of cancer, regulation in fatty acid metabolism, aspects of cardiovascular function, control of the central nervous system development, analysis of reproductive development and function are only some examples of research interests.

The final stage of cell differentiation is growth arrest. In animal tissues with rapid cell turnover terminally differentiated cells undergo programmed cell death. The cells have the ability to kill themselves by activating an intrinsic cell suicide program when they are no longer needed or have become seriously damaged. The execution of this program is termed apoptosis. Apoptosis is of importance for development and homeostasis of animals. The key components of this program have been conserved in evolution from worms (C. elegans) to insects (Drosophilia) to humans. The roles of apoptosis include the sculpting of structures during development, deletion of unneeded cells and tissues, regulation of growth and cell number, and the elimination of abnormal and potentially dangerous cells. In this way

apoptosis provides "quality control mechanism" that limits the accumulation of harmful cells, such as virus-infected cells and tumor cells. On the other hand inappropriate apoptosis is associated with a wide variety of diseases, including AIDS, neuro-degenerative disorders and ischemic stroke. Because it is now clear that apoptosis is a result of an active, gene-directed process, it should be eventually possible to manipulate this form of cell death by developing drugs that interact with its recently identified mechanisms of action. Inducers of cell differentiation, cell cycle arrest and apoptosis might be the novel molecular targets for new anticancer agents in addition to the signaling pathways for growth factors and cytokines.

Proteins, factors, receptors and genes of importance in apoptosis:

Proteases:

- Calpain, an intracellular cysteine protease, exact role unknown.

- Caspase-1 to Caspase-11, a family of proteases synthesized as an inactive proenzyme. Targets of the activated enzymes include: poly(ADP-ribose) polymerase, DNA-dependent protein kinase, U1 ribonucleoprotein, nuclear laminins and cytoskeleton components (actin).

- Granzyme B, a serine protease released by cytotoxic T-cells.

Receptors:

- CD 95 (synonyms: Fas, APO-1), a receptor protein of the TNF-receptor family which includes TNF-R1 and TNF-R2 with the common characteristic of a 70 amino acid cytoplasmic domain.

- FADD (synonym: MORT-1), a cytoplasmic protein

- DR-3 (synonym: APO-3) a member of the TNF-receptor-family

- DR-4 and DR-5

Genes:

- ced-3, ced-4 and ced-9 encode the general apoptotic and antiapoptotic program in Caenorhabditis elegans. Apaf-3 is the mammalian homologue of ced-3.

- Bcl-2 / Bcl-xL / Bax / Bcl-xS / Bak: a large gene family that can either inhibit or promote apoptosis.

- Cytokine response modifier A, a cowpox virus gene whose gene product inhibits caspases.

Others:

- Caspase-activated DNase (CAD) and its inhibitor (ICAD), causes DNA fragmentation in the nucleus

- Ceramide, a complex lipid that acts as a second messenger.

- c-Jun N-terminal kinase (JNK) is a proline-directed kinase

- p53 protein, is essential for the induction of apoptosis as a response to chromosomal damage.

- RAIDD, a death signal-transducing protein.

- Receptor interacting protein (RIP) is an accessory protein with a death domain and a serine/threonine kinase activity.

- Sphingomyelinase, an enzyme that hydrolyzes the complex lipid sphingomyelin to ceramide.

- Tumor necrosis factor (TNF) is a type -II membrane protein

- TNF-receptor associated factor (TRAF2), is an accessory protein that can bind to both TNF-R1 and TNF-R2.

Within the overall group of Differentiation/Development, several categories of proteins are coded for by clones of the invention:

Interleukins (e.g. Interleukin-7): Interleukin precursors related to interleukin-7, for example, are expected to act as new growth factors for human B lineage cells. Additionally,

these proteins should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells These interleukins could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells. (OMIN *146660). Clones in this category include: tes3_35e21.

Testis-specific Y-encoded proteins: The TSPY genes are arranged in clusters on the Y chromosome of many mammalian species. TSPY is believed to function in early spermatogenesis and is a candidate for GBY, the putative gonadoblastoma-inducing gene on the Y. Proteins of the TSPY-SET-NAP1L1 family represent proteins closely related to TSPY. These proteins seem to be involved in early spermatogenesis. Clones in this category include: fbr2_2d15.

Intracellular transport and trafficking

Eukaryotic cells rely for their viability on the partitioning of many basic cellular processes into membrane-bounded organelles. These are the nucleus, endoplasmic reticulum (ER), Golgi apparatus, endosomes, lysosomal compartments, mitochondria and peroxisomes. Most molecules destined for the lysosome, cell surface and outside the cell are routed through the ER and Golgi, which together with the vesicular intermediates between them, comprise the secretory pathway (Palade 1975). In the ER and Golgi compartments proteins are sorted, modified and often assembled into complexes *en route* to their final destination. Incorrectly assembled proteins are retained in the ER until they fold correctly or are targeted for degradation. Additional proteins are translocated into and function within the lumenal spaces of organelles or are secreted. Thus a large proportion of proteins synthesized require targeting to membranes either for insertion into or transport across them. A major purpose of this is growth. The secretory pathway is dependent on an intact cytoskeleton and also closely linked to general metabolism by affecting ribosome biogenesis (Mizuta and Warner, 1994). A huge number of proteins is required for targeting, translocation and sorting of newly synthesized proteins.

The first step in sorting is the recognition of cis-acting targeting or signal sequences that organelle-targeted proteins contain. This is carried out by cytosolic targeting factors and/or receptors on the membrane to which the protein is targeted. In some cases the primary

sequences are extremely degenerate, with only the overall character being conserved (hydrophobicity for an ER signal sequence, helical amphiphilicity for mitochondrial targeting sequence (Kaiser *et al.*, 1987; Lemire *et al.*, 1989). Following the targeting step, proteins are either inserted into or transported across the membrane (translocated) through a proteinaceous apparatus (termed the translocon). The translocon include or recruit motors to drive the translocation process in the correct direction (Schatz and Dobberstein, 1996).

Defined intracellular protein transport steps:

- ER
    - targeting to the ER
    - translocation into the lumen of the ER, and, depending on the presence of certain signals in the peptide sequence transport through the golgi complex
- Mitochondria
    - targeting
    - translocation
- Peroxisomes
- The general secretory pathway
    - protein modification, assembly and quality control in the ER
    - vesicle-mediated trafficking
    - vesicle docking and fusion
    - transport through the golgi apparatus and sorting at the trans-golgi
    - transport to the cell surface
    - transport routes to the lysosome
- Endocytosis
- Specialized protein transport routes
- Protein export from the cytoplasm

References: Palade, G (1975) Science 189:347-358; Mizuta et al. (1994) Mol Cell Biol 14: 2493-2502; Kaiser *et al.* (1987) Science 235: 312-317; Lemire *et al.* (1989) J Biol Chem 264: 20206-20215; Schatz et al. (1996) Science 271: 1519-1526.

Rab proteins

In eukaryotic cells the compartmentalisation of processes is a prerequisite for a tight regulation of processes and activities. The cells contain a highly dynamic set of membrane compartments that are responsible for packaging, sorting, secreting, and recycling proteins

and other molecules. Trafficking between organelles within the secretory pathway occurs as vesicles derived from a donor compartment fuse with specific acceptor membranes, resulting in the directional transfer of cargo molecules. This process is tightly controlled by the Rab/Ypt family of proteins (reviewed by Novick and Zerial, 1997), a branch of the superfamily of small GTPases. Rab proteins regulate a variety of functions, including vesicle translocation and docking at specific fusion sites. Rabs may also play critical roles in higher order processes such as modulating the levels of neurotransmitter release in neurons, a likely mechanism in synaptic plasticity that underlies learning and memory (Geppert and Südhof, 1998).

Small GTPases share a common three-dimensional fold that, in the GTP bound state, can bind a variety of downstream effector proteins. GTP hydrolysis leads to a conformational change in the "switch" regions that renders the GTPase unrecognizable to its effectors. In this way, by localizing and activating a select set of effectors, a common structural motif is used to control a wide array of distinct cellular processes.

The final steps in membrane fusion are likely to be driven by a set of proteins known as SNAREs. After a vesicle becomes docked, the cytoplasmic domains of VAMP (also termed synaptobrevin) and syntaxin on opposing membranes, in combination with a SNAP-25 molecule, coalesce into an elongated -helical bundle (Poirier et al., 1998 ; Sutton et al., 1998), which may lead to fusion. Because numerous SNARE isoforms have been identified that localize to distinct membrane compartments, it was originally proposed that the specificity of interaction between the SNARE proteins accounted for the specificity in membrane trafficking. Recent results, however, suggest that SNAREs are not specific in their ability to form complexes in vitro, suggesting that trafficking specificity requires additional factors (Yang et al., 1999). In this regard, Rab proteins are strong candidates for governing the specificity of vesicle trafficking. Like the SNAREs, many isoforms (40) of the Rab family have been identified that localize to specific membrane compartments (reviewed by Novick and Zerial, 1997).

Concomitant with the SNARE cycle, Rab proteins undergo a intricate cycle of membrane and protein interactions. Rabs are posttranslationally modified at C-terminal cysteines by the addition of two geranylgeranyl groups, which mediate membrane association when the Rab is in the GTP-bound state. After guanine nucleotide hydrolysis occurs, the Rab is extracted from the membrane upon forming a complex with a cytosolic GDP-dissociation

inhibitor (GDI). This cytosolic intermediate is then recycled onto a newly forming vesicle, most likely through a secondary factor termed a GDI dissociation factor (GDF), which displaces GDI. After the Rab becomes membrane bound, a guanidine nucleotide exchange factor (GEF) promotes release of GDP and the subsequent loading of GTP. In its GTP-bound conformation, the Rab is then free to associate with its specific set of effectors, which can in turn trigger events leading to the eventual fusion of the vesicle with a target membrane. To complete the cycle, perhaps after or concurrent with membrane fusion, a GTPase activating protein (GAP) accelerates nucleotide hydrolysis, switching off the GTPase. The remaining GDP-bound Rab can then participate in a new round of fusion.

Rab interactions with effectors are likely to regulate vesicle targeting and membrane fusion in three ways. First, a Rab may specifically facilitate vectorial vesicle transport. Vesicles are transported from their site of origin to acceptor compartments likely through associations with cytoskeletal elements and transport motors. A protein has been identified with a domain structure that suggests a connection between the cytoskeleton and the Rabs. This protein, called Rabkinesin-6, contains a kinesin-like ATPase motor domain followed by a coiled-coil stalk region and a RBD that specifically binds Rab6 (Echard et al., 1998). An additional link with the cytoskeleton is provided by the Rab effector, Rabphilin-3A. Rabphilin-3A has been shown in vitro to interact with -actinin, an actin-bundling protein, but only when not bound to Rab3A (Kato et al., 1996). These results raise the intriguing possibility that Rab proteins regulate vesicle interactions with the cytoskeleton and thereby play an active role in targeting vesicles to their appropriate destinations.

Second, Rab proteins may regulate membrane trafficking at the vesicle docking step. A number of Rab effectors, including Rabaptin-5, EEA1, Rabphilin-3A, and Rim, may serve as molecular tethers. Each effector protein contains a RBD, followed by a linker region (some having the potential to form elongated coiled-coil structures), and a domain capable of interacting with a second Rab or the target membrane. Rabaptin-5, for example, contains two RBDs, one near the N terminus that specifically recognizes Rab4 and a second near the C terminus that binds Rab5 (Vitale et al., 1998). Both Rim, which is localized to the target membrane, and Rabphilin-3A, which is localized to the vesicle, contain N-terminal RBDs and C-terminal Ca2+-binding C2 domains, implicating these effectors in synaptic vesicle localization or docking in response to Ca2+ influx (Wang et al., 1997). Tethering effectors may also recognize protein complexes on the acceptor membrane. Sec4p, a yeast Rab3A

homolog, interacts with the exocyst (Guo et al., 1999), a complex of seven or more subunits that is assembled at sites of vesicle fusion along the plasma membrane. The exocyst complex may therefore function as a landmark for Rab/effector-mediated vesicle docking.

Third, once a vesicle has become tethered to its fusion site, Rab proteins may selectively activate the SNARE fusion machinery. The mechanism of this activation is unknown but may involve direct interactions of Rabs or, more likely, their effectors with SNAREs. For example, Hrs-2 is a protein that binds to SNAP-25 and contains a Zn2+-finger motif characteristic of Rab-binding proteins such as Rabphilin-3A, Rim, EEA1, and Noc2, suggesting that Hrs-2 may form a physical link between Rabs and SNAREs (Bean et al., 1997). In addition, certain mutations in the syntaxin-binding protein Sly1p, the Sec1p homolog utilized in ER to Golgi trafficking, eliminate the requirement for Ypt1p, a Rab protein that functions at this trafficking step (Dascher et al., 1991). Rabs may therefore regulate SNARE associations through Sec1 family members. In support of this idea, a Rab effector was recently found to interact with a vacuole Rab, a Sec1p homolog, and a SNARE protein (Peterson et al., 1999), which suggests that this effector serves to connect Rab and SNARE function. In this way, Rabs and their effectors may facilitate the correct pairing of SNAREs.

References: Dascher et al. (1991) Mol. Cell. Biol. 11, 872-885; Echard et al. (1998). Science. 279, 580-585; Geppert et al. (1998) Annu. Rev. Neurosci. 21, 75-95; Guo et al. (1999). EMBO J. 18, 1071-1080; Kato et al. (1996) J. Biol. Chem. 271, 31775-31778; Novick et al. (1997) Curr. Opin. Cell Biol. 9, 496-504; Peterson (1999) Curr. Biol. 9, 159-162; Poirier et al. (1998) Nat. Struct. Biol. 5, 765-769; Vitale et al. (1998) EMBO J. 17, 1941-1951; Wang et al. (1997) Nature. 388, 593-598; Yang et al. (1999) J. Biol. Chem. 274, 5649-5653.

Within the overall group of Intracellular Transport and Trafficking several categories of proteins are coded for by clones of the invention.

Rab proteins:

Rab1B is essential for the intracellular transport of nascent low density lipoprotein (LDL) receptor. It is discussed as a universal mediator of endoplasmatic reticulum to Golgi transport of membrane glycoproteins in mammalian cells. . Clones in this category include: fbr2_2i17, fbr2_3b16.

Rab10 appear concentrated on membranes in the perinuclear region. Rab 10 has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases as reported by OMIN: 1) Choroideremia (OMIN *303199); and 2)RETT Syndrome (OMIN 312750). Clones in this category include: fbr2_62l19.

In mice, Rab17 shows epithelial cell specificity. Rab 17 is discussed as candidate gene for the mouse mutations ln (leaden), Tw (twirler), and ax (ataxia). Cloned from a brain cDNA library, the new putative Rab-protein is expected to be involved in vesicle trafficking within neuronal cells. These proteins can find application in modulating the transport of vesicles inside neuronal cells, which are essential for development of functional dendritic processes... Clones in this category include: fbr2_41m15.

Ankyrin G: The ankyrin 3 gene encodes a novel ankyrin, which is expressed in multiple tissues, with very high expression at the axonal initial segment and nodes of Ranvier of neurons in the central and peripheral nervous systems. Ankyrin G shows several tissue-specific alternative mRNA processing. The different ankyrin G proteins participate in maintenance/targeting of ion channels and cell adhesion molecules to nodes of Ranvier and axonal initial segments. Ankyrin G has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with Werner disease (OMIN *277700). Clones in this category include: fkd2_24p5.

Zn-T-transporters: The Zn-T-transporters are membrane proteins that facilitates sequestration of zinc in endosomal vesicles. In the brain, ZnT-3 mRNA seems to be involved in the accumulation of zinc in synaptic vesicles. Zinc (Zn) is an essential element in normal development and metabolism. Recent studies show that in Alzheimer's disease, Zn functions as a double-edged sword, affording protection against Alzheimer's amyloid beta peptide (the major component of senile plaques) at low concentrations and enhancing toxicity at high concentrations by accelerated aggregation of the amyloid beta peptide. These proteins can find application in modulation of Zinc transport in neuronal cells, thus providing means for a modulation of Alzheimer's amyloid beta peptide plaque formation. (OMIN *602878, *602095). Clones in this category include: fbr2_62f10.

Metabolism

This group includes proteins which are involved in the uptake and consumption of nutrients, and enzymes which are part of the biochemical pathways for energy metabolism or

which are involved in the supply of building blocks of nucleic acids, proteins (NTPs, dNTPs, amino acids) for DNA/RNA and protein synthesis, and fatty acids (membranes), to allow for the generation of higher order structures. This group constitutes the most important and largest group in prokaryotes and lower eukaryotes. The higher the evolutionary level of an organism is, however, the more other protein classes like 'signal transduction', 'cell cycle' and 'differentiation and development' increase in importance and number of representatives.

Proteins involved in the metabolism of energy and compounds (here: other than nucleic acids or proteins) are usually the products of house keeping genes, they are often constitutively and/or ubiquitously expressed.

Several categories of proteins are coded for by clones of the invention within the overall group of Metabolism:

NAT1, ARD1: In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1. NAT1 controls full repression of the silent mating type locus HML, sporulation and entry into G0. ARD1 is involved in the assembly of the NAT 1-complex. These can find application modulating NAT assembly and action and therefore could be important in metabolism of drugs and environmental mutagens.(OMIN *108345). Clones in this category include: fbr2_3g8.

Apolipoprotein E receptor: In LDL-receptors the class A domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of positively charged sequences in LDLR's ligands. These proteins can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins. In normal individuals, chylomicron remnants and very low density lipoprotein (VLDL) remnants are rapidly removed from the circulation by receptor-mediated endocytosis in the liver. In familial dysbetalipoproteinemia, or type III hyperlipoproteinemia (HLP III), increased plasma cholesterol and triglycerides are the consequence of impaired clearance of chylomicron and VLDL remnants because of a defect in apolipoprotein E. Accumulation of the remnants can result in xanthomatosis and premature coronary and/or peripheral vascular disease. OMIN reports that apolipoprotein has associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Familial hypercholesterolemia (OMIN 143890); 2) Familial combined hyperlipidemia (OMIN 144250); and 3) Alzheimer disease. (OMIN #104300). Clones in this category include: fbr2_62017.

Ubiquitin carboxyl-terminal hydrolases: Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquinated proteins. OMIN reports that Ubiquitin-specific proteases have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc…) with the following diseases: 1) Lung carcinoma (OMIN *603486); 2) x-linked retinal diseases (OMIN *300050); 3) oncogenesis (OMIN *300050);4) ovarian cancer (OMIN *300050). Clones in this category include: fbr2_78k24; htes3_27d1.

Phosphoserine signature (phosphoglucomutases, phosphomannomutase): These proteins take part in the conversion of hexose phosphates. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc…) with the following disease: Fanconi-Bickel Syndrome (OMIN #227810). Clones in this category include: fkd2_24b15.

NADH ubiquinone oxidoreductase: NADH:ubiquinone oxidoreductase is the first enzyme in the respiratory electron transport chain of mitochondria. It is a a membrane-bound multi-subunit protein. The bovine heart enzyme contains about 40 different polypeptides. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc…) with the following disease: Brancio-oto-renal syndrome (OMIN *6601445). Clones in this category include: fkd2_3o17.

Transketolases: Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc…) with the following diseases: Wernicke-Korsakoff Syndrome (OMIN *277730). Clones in this category include: tes3_17l17.

Fatty acid-CoA synthetases/ligases: These proteins contain AMP-binding domain signature(s), which is present in enzymes which act via an ATP-dependent covalent binding of AMP to their substrate. This domain is found in several CoA synthetases, such as acetate-CoA ligase (EC 6.2.1.1), long-chain-fatty-acid-CoA ligase (EC 6.2.1.3), bile acid-CoA ligase. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic,

causative, and/or related, etc...) with the following diseases: 1) Alport syndrome , mental retardation and elliptocytosis (OMIN *300157); 2) Adrenoleukodystrophy (OMIN *300100). Clones in this category include: tes3_35k17.

ADP/ATP or Adenine Nucleotide Translocataors: These proteins contain mitochondrial energy transfer signature(s) and are most abundant in mitochondria. In its functional state, it is a homodimer of 30-kD subunits embedded asymmetrically in the inner mitochondrial membrane. The dimer forms a gated pore through which ADP is moved from the matrix into the cytoplasm.. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) cardiomyopathy (OMIN *103220); 2) myopathy (OMIN *103220); 3)Progressive external ophthalmoplegia (OMIN *601227). Clones in this category include: tes3_35n12.

Carboxylesterases: OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1)hepatic carboxylesterase with detoxification of foreign compounds (OMIN *114835); 2) non-Hodgkin lymphoma (OMIN *114835); 3) B-cell chronic lymphocytic leukemia (OMIN *114835); 4) rheumatoid arthritis (OMIN *114835). Clones in this category include: tes3_35n9.

Heat shock proteins: OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1)27 kd heat shock protein has been correlated with thermotolerance in response to environmental challenges and developmental transitions. (OMIN *6021295). Clones in this category include: utel1_23e13.

Nucleic acid management

The genetic information is stored in the form of nucleic acids in all organisms. Two kinds of nucleic acids exist, DNA and RNA. Whereas the more stable DNA in most organisms constitutes the storage form of the genetic information, the labile RNA and in particular mRNA is an intermediate used for the temporal expression of specific genes.

In eukaryotes, DNA is usually a double stranded linear molecule consisting of two antiparallel strands and made up of a deoxyribose, a phosphorus backbone and the four bases A, C, G, and T. The DNA of some organisms has a ring structure. The structure of DNA was

unraveled years ago by Watson and Crick. DNA is directional molecule determined by the C-atoms of the sugar.

The most important processes dealing with nucleic acids are:

- replication (e.g. DNA polymerases, Telomerase)
- transcription (RNA polymerases)
- RNA processing (maturation - splicing and degradation)
- in addition, enzymes and proteins exist which require a nucleic acid (mostly RNA) in the active center to be functional (ribozymes - e.g. RNase, Ribosomal proteins)

The DNA of a cell is replicated in the S-phase of the cell cycle. Several enzymes carry out the task of doubling this nucleic acid. As all steps of the cell cycle, also the process of replication is tightly regulated. The enzyme DNA polymerase and several other proteins are involved in this process. Whereas many prokaryotes do have only one origin of replication (i.e., the starting point of the replication cycle), in eukaryotic DNAs (chromosomes) multiple such start points exist. The switch from the synthesis (S) phase to the subsequent G2 or M phases of the cell cycle are dependent on the completion of the replication. This makes clear, that a number of proteins are involved in the replication itself as well as in the control of the process. Since most eukaryotic chromosomes are linear structures, additional proteins and enzymes are necessary to make sure that the structure is maintained through successive generations. This includes those proteins necessary to build the three dimensional structure of chromosomes (e.g. histones) and the structural network of the nucleus and nucleolus (including the defined localization of transcriptionally active genes in the vicinity of nucleoli) but also such enzymes as telomerase which guarantees the integrity of the chromosomal ends.

The expression of genes is usually performed in two steps. First a messenger RNA (mRNA) is produced (transcribed) in one to many copies and second this mRNA is translated into the protein product. The regulation of transcription is discussed under the separate heading 'transcription factors', but also the classes 'signal transduction', 'development', 'cell cycle' and others are affected as the expression of certain genes determines the fate of a cell or organism.

The primary transcript (hnRNA - heterogeneous nuclear RNA) is a single stranded one-to-one copy of the gene as it is located on the chromosome. Before a protein can be translated, already during transcription the process of maturation is initiated. Firstly, a 5' cap structure is enzymatically and covalently added to the RNA, blocking the 5' end of the RNA.

Second, when the RNA polymerase has terminated polymerization, the enzyme poly A polymerase adds varying numbers of adenine residues to the 3' end of the transcript. This enzyme recognizes the sequence AAUAAA or AUUAAA (+ some minor variations), cuts the RNA 10 - 30 nucleotides downstream and adds the A residues. The size of the poly A sequence affects the stability of the RNA. Finally, in the process of splicing, the introns present on the genomic level and also present in the hnRNA are spliced out by a multi-protein complex consisting of several proteins and RNAs. The finally maturated mRNA is exported to the cytoplasm where it is translated with help of the ribozymes.

The half life of RNA is usually much shorter than that of DNA. Usually, the mRNA is degraded shortly after synthesis, to guarantee a very defined window of expression of a given gene. This regulation is necessary to specifically maintain or change the set of proteins present at any time in a cell. Specific regions in the 3'UTR (untranslated region) determine the stability of the mRNA in the cytoplasm before it is degraded by RNases, enzymes consisting both of protein and RNA.

References: Watson and Crick (1953) Nature **171**: 737-738.

Several categories of proteins are coded for by clones of the invention within the overall group of "Nucleic acid management"and include, among others, the following:

<u>RNA helicases including DEAD/H box helicases</u>: RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. DEAD box proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by with the following disease processes and/or genes: 1) ataxia-telangiectasia gene: "A human gene (DDX10) encoding a putative DEAD-box RNA helicase at 11q22-q23" *Genomics* 33:199-206, 1996, Savitsky et al., (OMIN *601235); 2) hematopoetic tumors: "Cloning and expression of a murine cDNA homologous to the human RCK/P54, a lymphoma-linked chromosomal breakpoint 11q23", Gene 166:293-6, 1995, Seto et al. (OMIN *600326); 3) dermatomyositis: a) "The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation."

Arthritis Rheum. 38: 1389-1399, 1995, Seelig et al. (OMIN *603277); b) "Two forms of the major antigenic protein of the dermatomyositis-specific Mi-2 autoantigen." (Letter), *Arthritis Rheum.* 39: 1769-1771, 1996., Seelig et al. (OMIN *603277); c) "The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities", *Cell* 95: 279-289, 1998. Zhang et al. (OMIN *603277); 4) Muscular Dystrophy, Pseudohypertrophic Progressive Duchenne and Becker Types (OMIN *310200); 5) Mucopolysaccharidosis Type IVA (OMIN *253000); 6) Albinism I (OMIN *203100); 7) Wilms Tumor 1 (OMIN *194070); 8) Spinocerebellar Ataxia 7 (OMIN *164500). Clones in this category include: fbr2_23b10, fbr2_3cl8, fbr2_6o17, fbr2_82i24, and tes3_14h21.

Inorganic pyrophosphatase: Inorganic pyrophosphatase (EC 3.6.1.1) (PPase) is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed as the product of the many biosynthetic reactions that utilize ATP. All known PPases require the presence of divalent metal cations, with magnesium conferring the highest activity. Clones in this category include: fbr2_64a15.

DNA-damage –inducible protein (dinP) or Proteins induced by DNA-Damage: The dinB/P pathway is a second SOS-pathway in E.coli. Genes related to this seem to be involved in modulating DNA repair and mutagenesis. Clones in this category include: fbr2_72b18.

Proteins with myc-type, helix-loop-helix dimerization domain signature(s). This helix-loop-helix domain mediates protein dimerization has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers. Therefore, these proteins could be novel DNA-binding proteins. Clones in this category include: fbr2_72l12.

Cytosolic ribosomal proteins L36: L36 seems to be part of the eukaryotic ribosomal peptidyl transferase center and can find application in modulation of ribosome assembly, maintenance and activity. Clones in this category include: fkd2_3b2.

Ribonuclease H: Ribonuclease H proteins are RNA modificating proteins and have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases as reported by OMIN: 1) Adenomatous Polyposis of the Colon (OMIN

*175100); 2) Retinoblastoma (OMIN *180200) ; and 3) Von Hippel-Lindau Syndrome (OMIN *193300). Clones in this category include: phtes3_15j3.

Signal transduction

Cells in higher order organisms need to continuously communicate with its environment especially with other cells of the same organism in order to maintain the function and specialization of the whole system these cells are part of. This important task of communication is performed with help of cell-surface receptors which receive and transmit signals from outside into the cell.

G-proteins

The largest known family of cell-surface receptors is that of the G-protein-coupled receptors, which mediate the transmission of diverse stimuli such as neurotransmitters, glycopeptides, hormones, peptides, odorant molecules, and photons. The functional unit of these receptors is composed of the receptor molecule itself (GPCR) which is anchored in the cytoplasma membrane with seven membrane spanning domains, the heterotrimeric G-protein which is composed of α and $\beta\gamma$-subunits ($G\alpha$ and $G\beta\gamma$), and the effectors that interact with $G\alpha$ and / or $G\beta\gamma$. In particular, the dissociated $G\alpha$ and $G\beta\gamma$ can regulate the activities of a number of effector molecules such as adenylate cyclases, phopholipase C isoforms, ion channels, and tyrosine kinases, resulting in a variety of cellular functions. The process of signal transduction must be tightly regulated and reversible in order to avoid overstimulation, to achieve signal termination, and render the receptor responsive to subsequent stimuli [Iacovelly L. et al., (1999) *FASEB J.* **13**, 1-8, Hamm, H.E. (1998) *J. Biol. Chem.* **273**, 669-672].

G-proteins are GTPases that, upon binding of GTP change their conformation which in return unmasks structural motives, in particular the so called effector loop, which can mediate the interactions to target proteins, or effectors, for the GTPases. This ability enables the GTPases to cycle between active, GTP-bound and inactive, GDP bound conformations and in the process to function as molecular traffic lights in a multitude of signal transduction pathways. The most important of these signal transduction pathways that are regulated with help of G-proteins are that of the phospholipase C / protein kinase C and that of the adenylate cyclase / protein kinase A.

The cycling of GTPases is tightly regulated by three main classes of proteins: The exchange of hydrolyzed GDP for a fresh GTP is facilitated by guanosine nucleotide exchange factors (GEFs), the hydrolysis of GTP to GDP is sped up by GTPase-activating proteins (GAPs), and the dissociation of GDP from the GTPases is inhibited by GDP dissociation inhibitors (GDIs) [Tapon and Hall (1997) *Curr.Opin. Cell. Biol.* **9**, 86-92, Van Aelst and D-Souza-Schorey (1997) *Genes Dev.* **11**, 2295-2322].

SOC-family

A conserved motif that was originally identified in proteins that negatively regulate the signaling action of cytokines was termed SOCS box, the Suppressor Of Cytokine Signaling. Based on homology, five distinct structural protein classes have been identified since that carry this motif. The function of most of these proteins is presently not known. Common to the proteins is only the SOCS box which is located near the C-terminus of the respective peptides. Recently, the SOCS box has been demonstrated to induce binding of proteins to elongins B and C which could target the proteins (and bound substrates) to the proteasomal protein degradation pathway (Kamura, T. *et al.* (1998) *Genes Dev.* **12**, 3872-3881; Zhang, J.-G. *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2071-2076).

The class where the SOCS box was originally described contains several members (SOCS-1-SOCS-7 and CIS). In addition to the SOCS box, these proteins also contain a SH2 (Src-homology 2) domain and a variable N-terminus. These SOCS proteins appear to form part of a classical negative feedback loop that regulates cytokine signal transduction. Upon cytokine stimulation, expression of SOCS proteins is rapidly induced and the proteins inhibit further cytokine action. The mode of action of the SOCS proteins is variable. While SOCS-1 binds and inhibits the JAK (Janus kinases) family of cytoplasmic protein kinases [Narahzaki M. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 13130-13134, Nicholson, S.E. *et al.* (1999) *EMBO. J.* **18**, 375-385], CIS appears to act by competing with signaling molecules such as the STATs (Transducers and Activators of Transcription) family for binding to phosphorylated receptor cytoplasmic domains [Yoshimura, A. *et al.* (1995) *EMBO J.* **14**, 2816-2826; Matsumoto, A. *et al.* (1997) *Blood* **89**, 3148-3154].

A second class of SOCS box protein contains additionally WD-40 repeats which were initially identified in the mouse WSB-1 and -2 proteins. The functions of WD-40 proteins are not completely understood but seem to be rather divergent. In Cdc4p the WD-40 repeats probably are necessary for binding the substrate for Cdc34p [Mathias, N. *et al.* (1999) *Mol.*

Cell Biol. **19**, 1759-1767]. Cdc4p is a component of a ubiquitin ligase that tethers the ubiquitin-conjugating enzyme Cdc34p to its substrates. The posttranslational modification of a protein by ubiquitin usually results in rapid degradation of the ubiquitinated protein by the proteasome. The transfer of ubiquitin to substrate is a multistep process where WD-40 repeats might play an important function.

Other WD-40 containing proteins (e.g. the retino blastoma binding protein RbAp48) have been shown to bind metal ions (Zinc) and that this metal binding might mediate and/or regulate protein-protein interactions which are functionally important in chromatin metabolism [Kenzior, A.L. and Folk, W.R. (1998) *FEBS Lett.* **440**, 425-429]. These proteins are involved in the RAS-cAMP pathway that regulates cellular growth [Ach R.A. *et al.* (1997) *Plant Cell* **9**, 1595-1606].

The SPRY domain has been identified in pyrin or marenostrin, a protein which is mutated in patients with Mediterranean fever and which is similar to the butyrophilin family. While butyrophilins seem to be involved in the lactation process in mammals, the function pyrin is unknown. Three proteins (SSB-1 to -3) have been identified to contain both SPRY and SOCS box motifs. The function of these proteins is also not known.

Ankyrin repeat containing proteins share a 33-residue repeating motif, an L-shaped structure with protruding β-hairpin tips which mediate specific macromolecular interactions with cytoskeletal, membrane, and regulatory proteins. These proteins play fundamental roles in diverse biological activities including growth and development, intracellular protein trafficking, the establishment and maintenance of cellular polarity, cell adhesion signal transduction, and mRNA transcription. Three proteins that contain ankyrin repeats (ASB-1 to -3) have been identified to contain a C-terminal SOCS box additionally to the ankyrin repeats. The function of these proteins or the individual domains remains to be discovered [Hilton, D.J. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 114-119].

A few small GTPases (RAR and RAR like) do also contain a SOCS box. GTPases are involved in signal transduction during cellular communication. The function of the SOCS box in this type of proteins is currently unclear [Hilton, D.J. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 114-119].

<u>Ca^{2+} as second messenger</u>

The bivalent cation Ca^{2+} is, besides cAMP, one of the two major second messengers in eukaryotic cells. Its intracellular concentration is tightly regulated and usually kept very

low compared to the cell's environment. Ca^{2+} binding proteins and transporters (Gap junction, Voltage-gated, second messenger-gated) help to sequester huge amounts of the ion in various organelles from where Ca^{2+} can be released upon extracellular stimuli. E.g. the contraction of the muscle is dependent on the presence of Ca^{2+} ions which are readily transported back into the organelles in order for the muscle to relax. In signal transduction, Ca^{2+} functions as a second messenger that activates Ca^{2+} dependent processes through the activation of Ca^{2+}/calmodulin dependent protein kinases (CaM kinases) which are the major effector molecules of Ca^{2+}. In the signaling cascades, the CaM dependent kinases activate phospholipases (e.g. phospholipase C) that in return activate other protein kinases such as protein kinase C.

cAMP

The cyclic AMP is produced by the enzyme adenylate cyclase in response to extracellular signals. Certain G-proteins stimulate the activity of adenylate cyclase which converts ATP to cAMP and PPi. Two molecules of cAMP bind to each of two regulatory subunits of cAMP dependent protein kinase which in turn dissociate from the two catalytic subunits of the heterotetramer R_2C_2. Upon release of the C-subunits, they become active and phosphorylate substrate proteins at Ser and Thr residues. The process leading from binding of extracellular molecules to their receptors, the transmission of the stimuli into the cell, the activation of adenylate cyclase and the subsequent activation of cAMP dependent protein kinase is one of two major signal transduction pathways in eukaryotic cells. Since the phosphorylation of proteins is a posttranslational modification of proteins, the kinases are described in the class "signal transduction."

SARA

Members of the transforming growth factor ß (TGFß) superfamily signal through a family of cell-surface transmembrane serine/threonine kinases, known as type I and type II receptors (Heldin et al., 1997 ; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). Ligand induces formation of heteromeric complexes of these receptors, and signaling is initiated when receptor I is phosphorylated and activated by the constitutively active kinase of receptor II (Wrana et al., 1994). The activated type I receptor kinase then propagates the signal to a family of intracellular signaling mediators known as Smads (contraction of the C.elegans Sma and Drosophila Mad genes which were the first identified members of this class of signaling effectors).

Three classes of Smads with distinct functions have been defined: the receptor-regulated Smads, which include Smad1, 2, 3, 5, and 8; the common mediator Smad, Smad4; and the antagonistic Smads, which include Smad6 and 7 (Heldin et al., 1997; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). Receptor-regulated Smads (R-Smads) act as direct substrates of specific type I receptors, and the proteins are phosphorylated on the last two serines at the carboxyl terminus within a highly conserved SSXS motif (Macías-Silva et al., 1996 ; Abdollah et al., 1997 ; Kretzschmar et al., 1997 ; Liu et al., 1997b ; Souchelnytskyi et al., 1997). Regulation of R-Smads by the receptor kinase provides an important level of specificity in this system. Thus, Smad2 and Smad3 are substrates of TGFß or activin receptors and mediate signaling by these ligands (Macías-Silva et al., 1996 ; Liu et al., 1997b ; Nakao et al., 1997), whereas Smad1, 5, and 8 are targets of BMP receptors and propagate BMP signals (Hoodless et al., 1996 ; Chen et al., 1997b ; Kretzschmar et al., 1997 ; Nishimura et al., 1998). Once phosphorylated, R-Smads associate with the common Smad, Smad4 (Lagna et al., 1996 ; Zhang et al., 1997), and mediate nuclear translocation of the heteromeric complex. In the nucleus, Smad complexes then activate specific genes through cooperative interactions with DNA and other DNA-binding proteins such as FAST1, FAST2, and Fos/Jun (Chen et al., 1996 , Chen et al., 1997a ; Liu et al., 1997a ; Labbé et al., 1998 ; Zhang et al., 1998 ; Zhou et al., 1998). In contrast to R-Smads and Smad4, the antagonistic Smads, Smad6 and 7, appear to function by blocking ligand-dependent signaling (reviewed in Heldin et al., 1997).

Phosphorylation of R-Smads by the type I receptor is essential for activating the TGFß signaling pathway (Heldin et al., 1997 ; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). However, little is known of how Smad interaction with receptors is controlled. A novel Smad2/Smad3 interacting protein has been described (Tsukazaki T. et al., 1998) that contains a double zinc finger, or FYVE domain, and which has been called SARA (Smad anchor for receptor activation). The SARA motif recruits Smad2 into distinct subcellular domains and co-localizes and interacts with TGFß receptors. TGFß signaling induces dissociation of Smad2 from SARA with concomitant formation of Smad2/Smad4 complexes and nuclear translocation. Moreover, deletion of the FYVE domain in SARA causes mislocalization of Smad2 and inhibits TGFß-dependent transcriptional responses. Thus, SARA defines a component of TGFß signaling that functions to recruit Smad2 to the receptor by controlling the subcellular localization of Smad.

References: Abdollah et al. (1997) J. Biol. Chem. 272, 27678-27685; Attisano et al. (1998) Curr. Opin. Cell Biol. 10, 188-194; Chen et al. (1996) Nature 383, 691-696; Chen et al. (1997a) Nature 389, 85-89; Chen et al. (1997b) Proc. Natl. Acad. Sci. USA 94, 12938-12943; Heldin et al. (1997) Nature 390, 465-471; Hoodless et al. (1996) Cell 85, 489-500; Kretzschmar et al. (1998) Curr. Opin. Genet. Dev. 8, 103-111; Kretzschmar et al. (1997) Genes Dev. 11, 984-995; Labbé et al. (1998) Mol. Cell 2, 109-120; Lagna et al. (1996) Nature 383, 832-836; Liu et al. (1997a) Genes Dev. 11, 3157-3167; Liu et al. (1997b) Proc. Natl. Acad. Sci. USA 94, 10669-10764; Macías-Silva et al. (1996) Cell 87, 1215-1224; Nakao et al. (1997) EMBO J. 16, 5353-5362; Nishimura et al. (1998) J. Biol. Chem. 273, 1872-1879; Souchelnytskyi et al. (1997) J. Biol. Chem. 272, 28107-28115; Tsukazaki et al. (1998) Cell 95, 779-791; Wrana et al. (1994) Nature 370, 341-347; Zhang et al. (1997) Curr. Biol. 7, 270-276; Zhang et al. (1998) Nature 394, 909-913; Zhou et al. (1998) Mol. Cell 2, 121-127.

Calcium

The bivalent cation Ca^{2+} is, along with cAMP, one of the two major second messengers in eukaryotic cells. Its intracellular concentration is tightly regulated and usually kept very low compared to the cell's environment. Ca^{2+} binding proteins and transporters (Gap junction, Voltage-gated, second messenger-gated) help to sequester huge amounts of the ion in various organelles from where Ca^{2+} can be released upon extracellular stimuli. E.g. the contraction of the muscle is dependent on the presence of Ca^{2+} ions which are readily transported back into the organelles in order for the muscle to relax. In signal transduction, Ca^{2+} functions as a second messenger that activates Ca^{2+} dependent processes through the activation of Ca^{2+}/calmodulin dependent protein kinases (CaM kinases) which are the major effector molecules of Ca^{2+}. In the signaling cascades, the CaM dependent kinases activate phospholipases (e.g. phospholipase C) that in return activate other protein kinases such as protein kinase C.

Rab proteins

In eukaryotic cells the compartmentalization of processes is a prerequisite for a tight regulation of processes and activities. The cells contain a highly dynamic set of membrane compartments that are responsible for packaging, sorting, secreting, and recycling proteins and other molecules. Trafficking between organelles within the secretory pathway occurs as

vesicles derived from a donor compartment fuse with specific acceptor membranes, resulting in the directional transfer of cargo molecules. This process is tightly controlled by the Rab/Ypt family of proteins (reviewed by Novick and Zerial, 1997), a branch of the superfamily of small GTPases. Rab proteins regulate a variety of functions, including vesicle translocation and docking at specific fusion sites. Rabs may also play critical roles in higher order processes such as modulating the levels of neurotransmitter release in neurons, a likely mechanism in synaptic plasticity that underlies learning and memory (Geppert and Südhof, 1998).

Small GTPases share a common three-dimensional fold that, in the GTP bound state, can bind a variety of downstream effector proteins. GTP hydrolysis leads to a conformational change in the "switch" regions that renders the GTPase unrecognizable to its effectors. In this way, by localizing and activating a select set of effectors, a common structural motif is used to control a wide array of distinct cellular processes.

The final steps in membrane fusion are likely to be driven by a set of proteins known as SNAREs. After a vesicle becomes docked, the cytoplasmic domains of VAMP (also termed synaptobrevin) and syntaxin on opposing membranes, in combination with a SNAP-25 molecule, coalesce into an elongated -helical bundle (Poirier et al., 1998 ; Sutton et al., 1998), which may lead to fusion. Because numerous SNARE isoforms have been identified that localize to distinct membrane compartments, it was originally proposed that the specificity of interaction between the SNARE proteins accounted for the specificity in membrane trafficking. Recent results, however, suggest that SNAREs are not specific in their ability to form complexes in vitro, suggesting that trafficking specificity requires additional factors (Yang et al., 1999). In this regard, Rab proteins are strong candidates for governing the specificity of vesicle trafficking. Like the SNAREs, many isoforms (40) of the Rab family have been identified that localize to specific membrane compartments (reviewed by Novick and Zerial, 1997).

Concomitant with the SNARE cycle, Rab proteins undergo a intricate cycle of membrane and protein interactions. Rabs are posttranslationally modified at C-terminal cysteines by the addition of two geranylgeranyl groups, which mediate membrane association when the Rab is in the GTP-bound state. After guanine nucleotide hydrolysis occurs, the Rab is extracted from the membrane upon forming a complex with a cytosolic GDP-dissociation inhibitor (GDI). This cytosolic intermediate is then recycled onto a newly forming vesicle,

most likely through a secondary factor termed a GDI dissociation factor (GDF), which displaces GDI. After the Rab becomes membrane bound, a guanidine nucleotide exchange factor (GEF) promotes release of GDP and the subsequent loading of GTP. In its GTP-bound conformation, the Rab is then free to associate with its specific set of effectors, which can in turn trigger events leading to the eventual fusion of the vesicle with a target membrane. To complete the cycle, perhaps after or concurrent with membrane fusion, a GTPase activating protein (GAP) accelerates nucleotide hydrolysis, switching off the GTPase. The remaining GDP-bound Rab can then participate in a new round of fusion.

Rab interactions with effectors are likely to regulate vesicle targeting and membrane fusion in three ways. First, a Rab may specifically facilitate vectorial vesicle transport. Vesicles are transported from their site of origin to acceptor compartments likely through associations with cytoskeletal elements and transport motors. A protein has been identified with a domain structure that suggests a connection between the cytoskeleton and the Rabs. This protein, called Rabkinesin-6, contains a kinesin-like ATPase motor domain followed by a coiled-coil stalk region and a RBD that specifically binds Rab6 (Echard et al., 1998). An additional link with the cytoskeleton is provided by the Rab effector, Rabphilin-3A. Rabphilin-3A has been shown in vitro to interact with -actinin, an actin-bundling protein, but only when not bound to Rab3A (Kato et al., 1996). These results raise the intriguing possibility that Rab proteins regulate vesicle interactions with the cytoskeleton and thereby play an active role in targeting vesicles to their appropriate destinations.

Second, Rab proteins may regulate membrane trafficking at the vesicle docking step. A number of Rab effectors, including Rabaptin-5, EEA1, Rabphilin-3A, and Rim, may serve as molecular tethers. Each effector protein contains a RBD, followed by a linker region (some having the potential to form elongated coiled-coil structures), and a domain capable of interacting with a second Rab or the target membrane. Rabaptin-5, for example, contains two RBDs, one near the N terminus that specifically recognizes Rab4 and a second near the C terminus that binds Rab5 (Vitale et al., 1998). Both Rim, which is localized to the target membrane, and Rabphilin-3A, which is localized to the vesicle, contain N-terminal RBDs and C-terminal Ca2+-binding C2 domains, implicating these effectors in synaptic vesicle localization or docking in response to Ca2+ influx (Wang et al., 1997). Tethering effectors may also recognize protein complexes on the acceptor membrane. Sec4p, a yeast Rab3A homolog, interacts with the exocyst (Guo et al., 1999), a complex of seven or more subunits

that is assembled at sites of vesicle fusion along the plasma membrane. The exocyst complex may therefore function as a landmark for Rab/effector-mediated vesicle docking.

Third, once a vesicle has become tethered to its fusion site, Rab proteins may selectively activate the SNARE fusion machinery. The mechanism of this activation is unknown but may involve direct interactions of Rabs or, more likely, their effectors with SNAREs. For example, Hrs-2 is a protein that binds to SNAP-25 and contains a Zn2+-finger motif characteristic of Rab-binding proteins such as Rabphilin-3A, Rim, EEA1, and Noc2, suggesting that Hrs-2 may form a physical link between Rabs and SNAREs (Bean et al., 1997). In addition, certain mutations in the syntaxin-binding protein Sly1p, the Sec1p homolog utilized in ER to Golgi trafficking, eliminate the requirement for Ypt1p, a Rab protein that functions at this trafficking step (Dascher et al., 1991). Rabs may therefore regulate SNARE associations through Sec1 family members. In support of this idea, a Rab effector was recently found to interact with a vacuole Rab, a Sec1p homolog, and a SNARE protein (Peterson et al., 1999), which suggests that this effector serves to connect Rab and SNARE function. In this way, Rabs and their effectors may facilitate the correct pairing of SNAREs.

References: Dascher et al. (1991). Mol. Cell. Biol. 11, 872-885; Echard et al. (1998). Science. 279, 580-585; Geppert et al. (1998). Annu. Rev. Neurosci. 21, 75-95; Guoet al. (1999). EMBO J. 18, 1071-1080; Kato et al. (1996). J. Biol. Chem. 271, 31775-31778; Novick et al. (1997). Curr. Opin. Cell Biol. 9, 496-504; Peterson et al. (1999). Curr. Biol. 9, 159-162; Poirier et al. (1998). Nat. Struct. Biol. 5, 765-769; Vitale et al. (1998). EMBO J. 17, 1941-1951; Wang et al. (1997). Nature. 388, 593-598; Yang et al. (1999). J. Biol. Chem. 274, 5649-5653.

Kinases

Reversible posttranslational modifications of proteins are major means of regulating cellular activities. Among the various modifications that are carried out by the cells, the addition of phosphoryl groups to Ser/Thr or Tyr residues is the most important and widely used. The phosphorylation of proteins is accomplished by protein kinases, while the reverse reaction, the removal of phosphoryl groups, is carried out by phosphatases. Kinases / Phosphatases regulate key positions e.g. in the processes of cell proliferation, differentiation and communication/signaling. These processes must be tightly regulated in order to maintain a steady state level of cellular fate. Mis-regulation of kinase activities (or that of

phosphatases) is made responsible for a multitude of disease processes such as oncogenesis, inflammatory processes, arteriosclerosis, and psoriasis.

Protein kinases constitute the largest protein family that is currently known. Several hundred kinases have been identified already. Classically, kinases are subdivided into two classes based on the amino acid residues in their substrates that are phosphorylated by the particular enzymes. The kinases specifically add phosphoryl groups from adenosine triphosphate (ATP) or, less frequently, guanosine triphosphate (GTP), either to serine and/or threonine or to tyrosine residues of substrate proteins. An estimated 1,000 to 10,000 proteins present in a typical mammalian cell are believed to be regulated also by the action of protein kinases.

Protein kinases are frequently integral parts of signaling cascades that transmit extracellular stimuli (e.g. hormones, neurotransmitters, growth- or differentiation factors) into the cell and result in various responses by the cells. The kinases play key roles in these cascades as they constitute a sort of 'molecular switches' turning on or off the activities of other enzymes and proteins, e.g. metabolic, regulatory, channels and pumps, receptors, cytoskeletal, transcription factors.

The regulation of kinase activities is accomplished by various means:

The best characterized example for the regulation via regulatory subunits is the cAMP-dependent protein kinase (PKA) which is also a prototype for second messenger activated protein kinases. This enzyme consists of a heterotetramer of two catalytic (C) and two regulatory (R) subunits. Upon binding of two molecules of second messenger (cAMP) in each R subunit, the catalytic subunits are released and active. Both of the catalytic and the regulatory subunits several isoforms exist. The combination of catalytic and regulatory subunits determines the localization of the holoenzyme and also the substrate spectrum that is available for phosphorylation. The consensus pattern necessary to be present in the substrate for PKA action is RRXS/T where X can be any amino acid.

The casein kinase II comprises another examples for holoenzymes that consist of catalytic and regulatory subunits. Other kinases that are activated by second messengers are cGMP-dependent protein kinase and Protein kinase C (PKC) which is activated by diacylglycerol, which in turn is produced by phospholipases by cleavage of phosphatidylcholine.

Receptor kinases usually consists of an extracellular domain which can bind effector molecules (e.g. growth factors and hormones) and transfer the stimulus to the intracellular domain of these proteins which usually is a protein tyrosine kinase. Other tyrosine kinases lack an extracellular domain but are associated with receptors which transfer the signal after effector binding by activating the associated protein kinase enzyme (e.g. Src kinase family; Src, Blk, Fgr, Fyn, Lck Lyn, Yes and Janus kinase family; Jak1-3, Tyk2).

Dysfunction of kinases, e.g. caused by non-functioning regulation, can be the cause of inflammatory diseases and uncontrolled proliferation. v-Src which is a truncated version of the C-Src protooncogene tyrosine kinase is a classical example for this process as v-Src does not contain the regulatory domain of the cellular gene and is thus constitutively active.

Several categories of proteins are coded for by clones of the invention within the overall group of "Signal transduction"and include, among others, the following:

Neurocalcin (Recoverin): Neurocalcin is a Ca(2+)-binding protein with three putative Ca(2+)-binding domains (EF-hands). In cattle, 6 isoforms are differentially expressed in the central nervous system, retina and adrenal gland. Homology with recoverin indicates involvement in Ca2+ dependent activation of guanylate cyclase.. These proteins can find application in modulating/blocking the guanylate cyclase-pathway. Diseases associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with these proteins include as reported by OMIN 1) autosomal dominant cone dystrophy (OMIN *600364); 2) cone dystrophy 3 (OMIN *600364); 3) cancer associated retinopathy (OMIN *179618). Clones in this category include: fbr2_23b21.

Proteins with a WW Domain: Proteins that contain a WW domain which has been originally described as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes. Examples of proteins containing the WW domain are Dystrophin, Utrophin, vertebrate YAP protein (binds the SH3 domain of the Yes oncoprotein), murine NEDD-4 (embryonic development and differentiation of the central nervous system), IQGAP (human GTPase activating protein acting on ras). Therefore these proteins should be involved in intracellular signal transduction. Diseases associated (as

potentially diagnostic, therapeutic, causative, and/or related, etc...) with these proteins include as reported by OMIN 1) Muscular Dystrophy, Pseudohypertrophic Progressive Duchenne and Becker Types (OMIN *310200). Clones in this category include: fbr2_23n16.

Protein substrates for cAMP-dependent protein kinase: Acting as a choride channel or chloride channel inhibitor these proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by OMIN with Cystic Fibrosis (OMIN #219700). Clones in this category include fbr2_82i17.

Sphingosine kinase: Sphingosine kinase is a new type of lipid kinase, which is regulated by growth factors. The enzyme phosphorylates sphingosine, which subsequently exerts intracellular and extracellular actions. Intracellulary, sphingosine 1-phosphate (SPP) promotes proliferation and inhibits apoptosis. In yeast, survival of cells exposed to heat shock indicates is dependent on SPP. Extracellulary, SPP inhibits cell motility and influences cell morphology, effects that appear to be mediated by the G protein-coupled receptor EDG1. These proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by OMIN with Gaucher Disease, Type I (OMIN *230800). Clones in this category include fbr2_82m6.

Vanilloid Receptors: VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. Related can find application as a target for the development of new nociception-modulating drugs. Clones in this category include tes3_20k2.

RCC1 (Regulator of chromosome condensation): RCC1 (regulator of chromosome condensation) is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein. RCC1 promotes the exchange of bound GDP with GTP, acting as a guanine-nucleotide dissociation stimulator. These proteins can find application in the regulation of gene expression by activition of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat. OMIN also reports that RCC1 has associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with retinitis pigmentosa (OMIN *312610). Clones in this category include tes3_21d4.

Ras inhibitor proteins: Ras is a signal transducting molecule involved in the receptor tyrosine kinase/RAS/Map kinase signalling cascade. Ras proteins bind GDP/GTP and show

intrinsic GTPase activity. Mutations in ras, which change aa 12, 13 or 61 activate the potential of ras to transform cultured cells and are implicated in a variety of human tumours. Ras inhibitor proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with many disease processes as reported by OMIN including: 1) Tumors of the lung, breast, brain, pituitary, pancrase, bone, skin, bladder, kidney, ovary, prostate and lymphocyte, Melanoma (OMIN *600160); 2) X-linked non-specific mental retardation (OMIN *300104); 3)adenomatouspolyposis of the colon (OMIN *175100); 4) Beckwith-Wieddemann Syndrome (#130650); and 5) Major affective disorder 1 (OMIN *125480). Clones in this category include ute1_22g21.

<u>Mammalian proteins cornicon involving the EGF-receptor</u>: Cornicon proteins are part of a signal transduction pathway involving the EGF-receptor. The EGF-receptor has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Familial hypercholesterolemia (OMIN 143890); 2) Leprechaunism (OMIN #246200); 3) Hemophilia B (OMIN *306900); 4) Ectodermal dysplasia 1; 5) Kartagenerer syndrome (OMIN *244400) and 6) Glioma of the brain (OMIN *137800).). Clones in this category include ute1_22e12.

<u>Transmembrane proteins</u>

Membrane region prediction was effected using the ALOM2 software (Klein et al., 1985; version 2 by K. Nakai). Similar to many other methods, the Kyte & Doolitle (1982) amino acid hydrophobicity scale is used in ALOM2 as the primary variable for classifying sequences in terms of their localization. High prediction accuracy is achieved through the system of intelligent decision rules and the utilization of a carefully selected training data set. The method also generates reliability estimates which makes it possible to distinguish between membrane-spanning proteins (I, intrinsic) and globular proteins with regions of high hydrophobicity buried in the core.

For a protein of length *L*, the block of length *l* with maximum hydrophobicity is found:

$$\max H = \max_{k=1,\ldots,L-l+1} (1/l) \sum_{i=k}^{k+l-1} H_i$$

where H_i represents the hydrophobicity of an individual residue.

Let P(I/maxH) and P(E/maxH) be the conditional probabilities that a protein is integral or peripheral, respectively, given its value of maximal hydrophobicity maxH, and let P(I) and P(E) be the prior probabilities of intrinsic and extrinsic membrane proteins estimated from the training set. Then a sequence is assigned to E if

P(E/maxH) > P(I/maxH)

or, after applying the Bayes rule,

P(E)P(maxH/E) > P(I)P(maxH/I),

where the conditional probabilities P(maxH/E) and P(maxH/I) can be determined based on the estimates of probability distributions of maxH in both groups.

Discriminant analysis allows to simplify this task by calculating the odds P(E/MaxH):P(I/maxH) as e^b, where b is the left-hand side of a linear or quadratic inequality. For example, for the window of length 17, the protein is allocated to the peripheral category E based on the empirically derived quadratic inequality:

$1.05(\text{maxH})^2+12.30\text{maxH}+17.49 > 0$,

whereas the optimal inequality for assigning membrane proteins (category I) is linear:

$-9.02\text{maxH} + 14.27 > 0$

The odds parameter can be made more or less stringent. For example, one can require odds at least 1:10 for a protein to be classified as integral. This leads to higher selectivity but less sensitivity.

The boundaries of membrane-spanning regions in putative membrane proteins are detected by means of an iterative procedure whereby the most hydrophobic region corresponding to the value maxH is considered to be membrane and removed from the sequence. The classification procedure is then repeated again for the remaining sequence, and, if such a protein is again classified as integral, the next most hydrophobic region is considered.

Reference: Klein, P., Kanehisa, M., DeLisi, C. (1985) The detection and classification of membrane-spanning proteins. *Biochem Biophys Acta* **815**: 468-476

Transcription factors

Purified eukaryotic RNA polymerase II is unable to initiate promoter-specific transcription. A family of factors that collectively confer RNAPII promoter specificity is known as the general transcription factors (GTFs). They include the TATA-binding Protein (TBP) TFIIB, TFIIE, TFIIF and TFI IH. These factors are conserved among all eukaryotes.

RNAPII complexes containing the entire set of GTFs or a subset of GTFs together with other proteins have been isolated from mammalian and yeast cells. Although purified RNAPII and GTFs are sufficient for promoter-specific initiation, this system fails to respond to activators. This is mediated by a further complex termed mediator complex which associates with the carboxy-terminal heptapeptide domain (CTD) of the largest subunit of RNAPII.

Purification of human RNAPII complexes resulted in two distinct forms of human RNAPII after analysis of functional properties. One complex contained chromatin remodeling activities but was devoid of GTFs. The other complex did not contain factors that modify chromatin but contained a subset of SRB/mediator subunits and GTFs and other polypeptides that mediate transcriptional activation, a scenario similar to that reported for yeast.

A complex designated NAT (~2O SU) for negative regulator of transcription contains RNAPII, Cdk8, homologs of the yeast mediator complex as well as Rgrl and Srb1O/11 known as negative regulators of transcription.

A complex with striking similar structural and functional properties to NAT has been identified designated SMCC (~I5 SU) (SRB/mediator coactivator complex), that can also mediate transcriptional activation.

The SMCC complex includes all reported NAT subunits including subunits of the TRAP complex. TRAP is a coactivator complex isolated on the basis of its interaction with the thyroid hormone receptor. Another coactivator complex DRIP, isolated on the basis of its

ability to interact with the vitamin D3 receptor, contains novel subunits as well as subunits of NAT/SMCC and TRAP complexes.

The effects of each of these coactivator complexes is dependent on the TFIID complex. It is not known if the T AF subunits of TFIID are required. It is likely that new coactivator complexes will be uncovered containing both novel and previously defined components.

Beside the huge amount of transcription factors which can be part of the RNAIIP holoenzyme or the coactivator complexes there is an even larger quantity of specific transcription factors binding to promoter elements within the DNA sequences of a given gene leading to activation or repression of transcription. A broad range of cellular responses like differentiation, proliferation, cell death and others are elicited through activating or repressing the transcription of target genes.

There are at least five superclasses of transcription factors:

1. Superclass contains members with characteristic basic domains:

Members are:

Leucine zipper factors, where the basic domain is followed by a leucine zipper of repeated leucine residues at every seventh position. The zipper mediates protein dimerization as a prerequisite for DNA-binding.

Helix-loop-helix factors (bHLH) contain a DNA-binding basic region followed by a motif of two potential amphipathic alpha-helices connected by a loop of variable length also mediating dimerization.

Factors with a combination of Helix-loop-helix and leucine zipper.

Further members of this superclass are NF-1, RF-X, and bHSH like proteins.

2. Superclass comprises factors containing zinc-coordinating DNA-binding domains.

Members are:

Proteins with Cys4 zinc finger of nuclear receptor type, where two such motifs differing in size, composition and function are present in each receptor molecule. Each finger comprises 4 cysteine residues coordinating one zinc ion. The second half including the second cysteine pair has alpha-helix conformation and the helix of the first finger binds to the DNA through the major groove. The sequence between the first two cysteines of the second finger mediates dimerization upon DNA-binding. This class includes the steroid hormone receptors and the thyroid hormone receptor-like factors. Other diverse cys4 zinc fingers have a motif of GATA-type.

Proteins with Cys2His2 zinc finger domain(s). Each finger comprises 2 cysteine and 2 histidine residues coordinating one zinc ion, and in some cases one histidine is replaced by another cysteine. The zinc ion is essential for DNA-binding.

Proteins with Cys6 cysteine-zinc cluster(s). Six cysteine residues coordinate two zinc ions, i. e. two of the thiol groups are coordinating two zinc ions each. Present in many fungal regulators.

Zinc fingers of alternating composition.

3. Superclass contains factors of helix-turn-helix type.

Members are:

Proteins with homeo domains. Homeo domains are three consecutive alpha-helix structures. Helix 3 contacts mainly the major groove of the DNA, some contacts at the minor groove are observed as well. Helix 2 and 3 resemble the helix-turn-helix structure of prokaryotic regulators.

Proteins with Paired box domain(s). This is a DNA-binding domain of approximately 130 amino acid residues. Its N-terminal half is basic, its C-terminal half is highly charged in general. It probably comprises 3 alpha-helices.

Proteins with Fork head / winged helix domain(s). This domain was identified by homology between HNF-3A and fkh. The domain comprises approx. 110 AA. Analysis of the crystal structure has revealed a compact structure of three alpha-helices, the third alpha-helix

being exposed towards the major groove of the DNA. The domain also exerts minor groove contacts. Upon binding to DNA, it induces a bend of 13 degree.

Heat shock factors

Proteins with Tryptophan clusters. The tryptophan clusters comprise several tryptophan residues with a spacing of 12-21 amino acid residues; the subclass of myb-type DNA-binding domains typically exhibit a spacing of 19-21 amino acid residues.

Proteins with TEA domain(s). The TEA domain has been identified as a region which is conserved among the transcription factors TEF-l, TECl and abaA. This domain in TEF-l has been shown to interact with DNA, although two additional regions may also contribute to DNA-binding. lt is predicted to fold into three alpha-helices, with a randomly coiled region of 16-18 amino acid residues between helices 1 and 2, and a short stretch between helices 2 and 3 of 3-8 residues.

4. Superclass contains beta-Scaffold Factors with Minor Groove Contacts

Members are:

Proteins with RHR (Rel homology) region.

The structure of the Rel-type DBD exhibits a bipartite subdomain structure, each subdomain comprising a beta-barrel with five loops that form an extensive contact surface to the major groove of the DNA. Particularly, the first loop of the N-terminal subdomain (the highly conserved recognition loop) performs contacts with the recognition element on the DNA, but other loops are involved. The fact that the main DNA-contacts are made through loops has been suggested to provide a high degree of flexibility in binding to a range of different target sequences. Augmenting interactions are achieved by two alpha-helices within the N-terminal Part that form strong minor groove contacts to the A/T-rich center of the B-element. In p65, the sequence between both alpha-helices is much shorter and even helix 2 is truncated. The second, C-terminal domain is necessary mainly for protein dimerization.

p53 proteins

MADS (MCMl-agamous-deficiens-SRF) box proteins. Proteins of this class comprise a region of homology. The DNA-binding domain also comprises the dimerization capability. In the DNA-bound dimer (shown for SRF), two antiparallel amphipathic alpha-helices (alpha-I), form a coiled coil and are oriented approximately parallel on the minor groove. These helices make minor and major groove contacts, the N-terminal extensions form minor groove contacts. The bound DNA is bent and wrapped around the protein. It exhibits a compressed minor groove in the center and widened minor groove in the flanks.

Beta-Barrel alpha-helix transcription factors.

TATA-binding proteins

HMG proteins

Proteins of this class comprise a region of homology with the chromosomal non-histone HMG proteins such as HMG1. This region comprises the DNA-binding domain which in some instances such as HMG1 mediates sequence-unspecific, in other cases such LEF-1 sequence-specific binding to DNA. This domain exhibits a typical L-shaped conformation made up of 3 alpha-helices and an extended N-terminal extension of the first helix. The latter together with helix 1, which contains a kink, form the long arm of the L, whereas helices 1 and 2 form the short arm. Binding to the minor groove induces a sharp bending of the DNA by more than 90 degree, away from the bound protein. The overall topology of the DNA-protein complexes resembles somewhat that of the TBP-TATA box complex.

Heteromeric CCAAT factors

Proteins with Grainyhead domain(s)

Cold-shock domain factors. Cold-shock domain proteins are characterized by a highly conserved region first found in prokaryotic cold-shock proteins. This domain is a single-stranded nucleic acid-binding structure interacting with DNA or RNA. It consists of an antiparallel five-stranded beta-barrel, the strands of which are connected by turns and loops. Within this structure, a three-stranded beta-strand contains a conserved RNA-binding motif, RNPl. Not all CSD proteins are transcription factors. Those which specifically bind to a

certain sequence are termed Y-box proteins. Proteins of this class were previously called protamine-like domain proteins because of having a highly positively charged domain with interspersed proline residues.

Proteins with Runt homology domain

The members of this transcription factor class have been identified on the basis of their homology to a defined region within the Drosophilia protein Runt. The runt domain is part of the DNA-binding domain of these factors. It consists mainly of beta-strands, does not contain alpha-helical regions and seems to be most similar to the palm domain found in DNA polymerase beta (rat).

5. Superclass contains other transcription factors like Copper fist proteins, HMGI(Y), STAT, Pocket domain proteins and Ap2/EREBP-related factors.

The classification of transcription factors originates from TRANSFAC database:

http: //transfac.gbf.de/TRANSFAC/

Reference: Heinemeyer

Several categories of proteins are coded for by clones of the invention within the overall group of "Transcription Factors".and include, among others, the following:

Dcoh: Dcoh is a bifunctional protein, complexed with biopterin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the biopterin cofactor of phenylalanine hydroxylase. The Dcoh protein has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) hyperphenylalanemia (OMIN 126090, #264070). Clones in this category include fkd2_46k12.

Signal transducing proteins: Beta-transducin subunits of G-proteins contain WD-40 repeats. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. Due to the zinc finger the novel protein seems to be a new molecule involved in signal transduction and transcription. These proteins have been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) essential hypertension (OMIN *139130). Clones in this category include ute1_1i2.

* * *

The invention, therefore, specifically contemplates the following assemblages of materials, which track the above-identified fourteen functional groupings, that are useful in practicing the profiling aspects of the invention. One type of assemblage is nucleic acid-based and can include the following groupings of sequences and their derivatives: all sequences; human fetal brain sequences; brain derived sequences; human fetal kidney library sequences; kidney derived sequences; human mammary carcinoma library sequences; mammary carcinoma derived sequences; human testis library sequences; testes derived sequences; cell cycle genes; cell structure and motility genes; differentiation and development genes; intracellular transport and trafficking genes; metabolism genes; nucleic acid management genes; signal transduction genes; transmembrane protein genes; and transcription factor genes. Other assemblages contain proteins or their corresponding antibodies or antibody fragments, divided along the same groupings.

Database Applications

Because they are human genes and gene products, the inventive molecules are useful as members of a database. Such a database may be used, for example, in drug discovery and rationale drug design or in testing the novelty and non-obviousness of newly sequenced materials. In addition, they are particularly suited in designing variants for the profiling (and other) applications described herein. Hence, the following discussion of electronic embodiments applies equally to such variants, which, naturally, will be generated and stored using a computer using known methodologies.

Accordingly, one aspect of the invention contemplates a database of at least one of the inventive sequences stored on computer readable media. Again, the individual sequences may be grouped with regard to the individual functional and structural groups mentioned above. While the individual sequences of a database may exist in printed form, they are preferably in electronic form, as in an ascii or a text file. They may also exist as word processing files or they may be stored in database applications like DB2, Sybase, Oracle, GCG and GenBank. One skilled in the art will understand the range of applications suitable for using and storing the electronic embodiments of the invention.

"Computer readable media" refers to any medium which can be read and accessed by a computer. These include: magnetic storage media, like floppy discs, hard drives and magnetic tape; optical storage media, like CD-ROM; electrical storage media, like RAM

and ROM; and hybrids of these categories, like magnetic/optical storage media. One skilled in the art will readily understand the scope of computer readable media and how to implement them.

Biological Activities and Assays for Implementing Therapeutic and Diagnostic Applications

This section provides assays for biological activity that are useful in characterizing and quantifying the biological activity of the inventive molecules and their derivatives, which is relevant to the pharmacological effects of the inventive molecules. As used in this section, it will be understood that "protein" may also refer to the inventive antibodies (including fragments).

Cytokine and Cell Proliferation/Differentiation Activity

A protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M + (preB M +), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin gamma , Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6-Nordan, R. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11-Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9-Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

Immune Stimulating or Suppressing Activity

A protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by vital (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to modify immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the

tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function in vivo on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-vital immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient.

The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected ex vivo with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection in vivo.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and beta 2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Bowmanet al., J. Virology 61:1992-1998; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Thl/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of

Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

Hematopoiesis Regulating Activity

A protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

Tissue Growth Activity

A protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the

treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendonitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and

cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

Activin/Inhibin Activity

A protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle

stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin alpha family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin- beta group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

Chemotactic/Chemokinetic Activity

A protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of

cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

Hemostatic and Thrombolytic Activity

A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

Receptor/Ligand Activity

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such

receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in:Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

Anti-Inflammatory Activity

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of

cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

Tumor Inhibition Activity

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

Other Activities

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or caricadic cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in

a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

Particular Applications for Certain Clones

The following sets out a non-exclusive list of applications for certain embodiments of the invention. In the interest of economy, applications relevant to multiple embodiments are not duplicated in this list. Other embodiments described in below have similar characteristics, as described therein. The artisan is directed, therefore, to this section for similar descriptions of the functions of other embodiment.

Testes

htes3_15c24: The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

htes3_15i5: The new protein can find application in modulating the structure of the human spermatozoa radia spoke head and modulation of sperm motility in men.

htes3_15k11: The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

htes3_17n12: The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

htes3_20k2: The new protein can find application as a target for the development of new nociception-modulating drugs.

htes3_20m18: The new protein can find application in modulation of mitochondrial DNA replication and maintenance.

htes3_20d4: The new protein can find application in the regulation of gene expression by activition of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat.

htes3_21j15: NY-CO-33 is a protein recognised by autologous antibodies of human colon cancer patients. The novel protein contains 4 C2H2 Zinc fingers and is a new putativ transcription factor. The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division.

htes3_26g22: The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division. The novel TBP-binding protein is considered to participate in transcription regulation through the interaction with TBP. The new protein can find application in modulation of gene transcription.

htes3_21l16: The new protein can find application in modulation of protein translocation into the endoplasmic reticulum.

htes3_27d1: The novel protein can find application in modulation of ubiquitin- and protein metabolism in cells.

htes3_2m18: The novel protein can find application as multifunctional nuclease / exoribonuclease.

htes3_35b4: The new protein can find application in modulation of the mitotic spindle.

htes3_35b5: The novel protein can find application in modulating the v-ATPase activity in endocytic and secretory organelles.

htes3_35e21: Due to the close relationship to human interleukin-7, the novel interleukin is expected to act as a new growth factor for human B lineage cells. Additionally, the protein should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells. This new interleukin could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells.

htes3_35k16: Therefore it is a new fatty acid-CoA synthetasese/ligase with unknown substrate. The new protein can find application in modulation of fatty acid metabolism and as a new enzyme for biotechnologic production processes.

htes3_35n12: The new protein can find application in modulation of ADP-transport and energy metabolism in cells/mitochondria.

htes3_35n9: The new protein can find application in modulation of carboxylester metabolism and as a new enzyme for biotechnologic production processes.

htes3_35p22: The novel protein is closely raleted to human tre-2 and other enzymes involved in the degradation of ubiquitinated proteins. The human tre-2 oncogene encodes a deubiquitinating enzyme, indicating a role for the ubiquitin system in mammalian growth control. The novel protein can find application in cancer diagnostics and treatment, and in regulating protein stability and growth control via regulation of ubiquitination.

htes3_4h6: The novel kinesin protein can find application in modulating the function of kinesin and modulating intracellular transport via/on microtubules.

htes3_72k15: FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin-binding protein. The gene locus fgd1 seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an esin yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. The novel protein seems to be the human orthologue of rat frabin.

The new protein can find application in modulating of cell structure and motility as well as modulation of the JNK/SAPK pathway.

htes3_72p16: As Mem3, the novel protein is similar to yeast VPS (vacuolar protein sorting) 35. The null allele of VPS35 results in yeast in a differential defect in the sorting of vacuolar carboxypeptidase Y (CPY), proteinase A (PrA), proteinase B (PrB), and alkaline phosphatase (ALP). The new protein can find application in modulation the sorting of proteins into different compartments.

htes3_7b22: The novel protein is related to paramyosin, a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as Schistosoma mansoni. The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamic.

htes3_7j3: The new protein is closely related to C-Tak1 and therefore should be involved in cell-cycle regulation, too. The new protein can find application in modulating/blocking the cell cycle.

htes3_7p9: The nuclear domain (ND)10 also described as POD or Kr bodies is involved in the development of acute promyelocytic leukemia and virus-host interactions. The NDP52 protein is part of this complex structure. In vivo, NDP52 is transcribed in all human tissues, but is redistributed upon viral infection and interferon treatment. ND10 plays an important role in the viral life cycle. The novel protein is similar to NDP52. It contains three leucine zippers and a RGD cell attachment site. This protein seems to be a novel part of the ND819) complex. The new protein can find application in modulation of viral infections and tumour events.

htes3_8m10: The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA. The new protein can find application in modulation of mRNA translation and processing/stability.

Kidney

hfkd2_24b15: The new protein can find application in modulation of hexose metabolism pathways and as a new enzyme for biotechnologic production processes.

hfkd2_24n20: The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton. The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamics.

hfkd2_3o17: The new protein can find application in modulation of the respiratory electron transport chain pathways of mitochondria.

hfkd2_46j20: The new protein can find application in modulating the homoprotocatechuate degradative pathway and as a enzyme for biotechnologic production processes.

hfkd2_46k19: The new protein can find application in modulating/blocking the expression of genes controlled by the hepatocyte nuclear factor-1.

hfkd2_46m4: SAR1 proteins are involved in vesicular transport between the endoplasmic reticulum and the Golgi apparatus.

hfkd2_46k14: rab6 is a ubiquitous ras-like GTPase involved in intra-Golgi transport. The new protein can find application in modulating the transport of vesicles inside the Golgi apparatus.

Uterus Associated:

hutel_18i19: The SREBP-2 protein is embedded in the membranes of the nucleus and endoplasmic reticulum. In cholesterol-depleted cells the proteins are cleaved to release soluble NH2-terminal fragments that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis. The new protein is a putative transcription factor capable of protein-protein interaction via a lim domain and additionally shows similarity to the common sunflower transcription factor SF3.

hutel_18l1: The novel protein is similar to several 40S ribosomal proteins and therefore seems to part of the corresponding ribosome sub-unit.

hutel_19g22: The new protein can find application in modulation of tissue-calcification, especially the uterus.

hutel_19h17: The new protein can find application in modulating the response of cells to oxysterols.

hutel_20b19: The novel protein seems to be a novel enzyme with sarcosine oxidase activity. The new protein can find application in modulation of sarcosine metabolism and as a new enzyme for biotechnologic production processes.

hutel_20g21: The novel protein seems to be a new ras inhibitor protein. The new protein can find application in modulating/blocking ras dependent signal transduction pathways.

hutel_20h13: The novel protein is a new human alpha-adaptin. The new protein can find application in modulating endocytosis and vesicle trafficking in cells.

hutel_20m11: The new protein can find application in modulating/blocking the activity of protein phosphatase-1 and in modulating the cell cycle.

hutel_20m24: This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2. The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

hutel_22e12: The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

hutel_23e13: The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnologic production processes.

hutel_24j6: The new protein can find application in modulation of cell-cell-adhesion.

hutel_24h3: The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

Fetal Brain:

hfbr2_16c16: The new protein can find application in modulating/blocking of cyto skeleton-membrane protein interaction.

hfbr2_23b21: The new protein can find application in modulating/blocking the guanylate cyclase-pathway.

hfbr2_23b10: The new protein can find application in modulation of splicing.

hfbr2_2b5: The novel protein contains the typical (xxG)n repeat of collagen proteins and a Pfam von Willebrand factor type A domain. Therefore, the protein seems to be a new collagen alpha chain. The new protein can find application in modulation of connective tissue, bone and cartilage development and maintainance.

hfbr2_2c17: The new protein can find application in modulating/blocking G-protein-dependent pathways.

hfbr2_2d15: The new protein can find application in modulating early spermatogenesis.

hfbr2_2i17: The new protein can find clinical application in modulating the transport of glycoproteins inside cells, especially of the LDL receptor.

hfbr2_2k14: Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. In addition, the novel protein contains a RGD cell attachment site. Therefore the novel protein is a new putative tumour-suppressor gene.

hfbr_3c18: RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and is a new member of this subgroup.

hfbr_3g8: The new protein can find application modulating NAT assembly and action and therefore be important in metabolism of drugs and environmental mutagens.

hfbr2_62b11: The rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

hfbr2_62o17: The new protein can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins.

hfbr_6b24: The new protein can find application in modulation of rhamnose metabolism and as a new enzyme for biotechnologic production processes.

hfbr_72b18: The new protein can find application in modulating DNA repair and mutagenesis.

hfbr_78c4: The new protein can find application in modulating/blocking the response of cells to interferons.

hfbr_78k24: These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquinated proteins. The new protein can find application in modulation of protein stability/degradation in cells.

hfbr_82e4: The new protein can find clinical application in modulating/blocking calmodulin-mediated pathways in human neuronal cells.

VARIANTS OF THE INVENTIVE DNA MOLECULES

Variants in General

"Variants," according to the invention, include DNA and/or protein molecules that resemble, structurally and/or functionally, those set forth in herein. Variants may be isolated from natural sources ("homologs"), may be entirely synthetic or may be based in part on both natural and synthetic approaches.

The section set forth below presents various structural and functional characteristics of molecules within the invention. Preferred molecules are characterized by a combination of one or more of these characteristics. For instance, some preferred molecules are described with reference to at least two structural characteristics, while others may be described with reference to at least one structural and at least one functional characteristic.

It will be recognized by the skilled artisan that structure ultimately defines function, *i.e.* the functions of the molecules described herein derives from the structures of those

molecules. Accordingly, the structural variants described below that bear the closest structural relationship (as variously defined below) to the inventive molecules are the variants that most likely will preserve biological function. This relationship between structure and function will guide the skilled artisan in identifying the preferred embodiments of the invention.

Splicing Variants

It is well-known that eukaryotic structural genes are comprised of both protein coding and non-coding portions. When the messenger RNA is transcribed from the DNA template, it contains introns, which are non-coding, and exons, which are coding. In order to form a translation competent mRNA, the introns must be "spliced" out of this initial pre mRNA.

Specific sequences within the pre mRNA represent "splice junctions" that direct the cellular splicing machinery to the appropriate position. The splice junctions are loosely conserved sequence regions of the pre mRNA, which almost invariably begin with GT and end with AG (DNA perspective). The 5' end of the splice junction typically contains about nine somewhat conserved residues, for example, C/AAGTA/GAGT. The 3' end usually contains a pyrimidine rich stretch of at least about 11 nucleotides, followed by NC/TAGG. Splicing occurs before the GT and after the AG. Mount, *Nucleic Acids Res.* 10:459-72 (1982).

Interestingly, exons often correspond to discrete functional domains of the protein product. The intron/exon arrangement thus creates a linear array of nucleotides which can be correlated to discrete, and often interchangeable, functional protein fragments. Go, *Nature* 291:90-92 (1981); Branden *et al.*, *EMBO J.* 3:1307-10 (1984). This linear arrangement creates the possibility of generating multiple different full length proteins by rearranging the order of the different functional portions in the array. For example, if a set of exons are arranged 1-2-3-4, where (-) represents the introns separating the exons, a splicing event need not simply produce 1234, but may produce 123, 134, 124 and so on. Production of different mRNA products in this way is commonly called "alternative splicing." Andreadis *et al.*, *Ann. Rev. Cell Biol.* 3:207-42 (1987).

Some of the present DNA molecules can be represented in modular fashion in terms of their coding regions. Essentially, these modules are exons (though each "exon" may in fact be made up of several exons), which may be combined in different ways to form a variety of

different DNA molecules, each encoding a different functional protein. Splicing variants are indicated below.

Degenerate Variants

One aspect of the present invention provides "degenerate variants" of the nucleic acid fragments of the present invention. A "degenerate variant" is a nucleotide fragment which differs from those of inventive molecules by nucleotide sequence, but due to the degeneracy of the genetic code, encodes an identical polypeptide sequence.

Given the known relationship between DNA sequences and the proteins they encode, degenerate variants typically are described by reference to this relationship. It is well known that the degeneracy of the genetic code results in many possible DNA sequences which encode a particular protein. Indeed, of the three bases which comprise an amino acid-encoding triplet, the third position, and often the second, almost always may vary. This fact alone allows for a class of variant DNA molecules which encode protein sequences identical to those disclosed herein, yet have about 30% sequence variation. In other words, the variant DNA molecules are about 70% identical to the inventive DNAs, having no additional or deleted sequences. Thus, one aspect of the invention provides degenerate variant DNA molecules encoding the inventive protein sequences.

In one embodiment, these variants have at least about 70% sequence identity with the DNA molecules described herein. In a preferred embodiment, these variants have at least about 80% sequence identity to the inventive molecules. In a more preferred embodiment these variants have at least about 90% sequence identity with the inventive molecules.

Conservative Amino Acid Variants

Variants according to the invention also may be made that conserve the overall molecular structure of the encoded proteins. Given the properties of the individual amino acids comprising the disclosed protein products, some rational substitutions will be recognized by the skilled worker. Amino acid substitutions, *i.e.* "conservative substitutions," may be made, for instance, on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved.

For example: (a) nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; (b) polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine;

(c) positively charged (basic) amino acids include arginine, lysine, and histidine; and (d) negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Substitutions typically may be made within groups (a)-(d). In addition, glycine and proline may be substituted for one another based on their ability to disrupt α-helices. Similarly, certain amino acids, such as alanine, cysteine, leucine, methionine, glutamic acid, glutamine, histidine and lysine are more commonly found in α-helices, while valine, isoleucine, phenylalanine, tyrosine, tryptophan and threonine are more commonly found in β-pleated sheets. Glycine, serine, aspartic acid, asparagine, and proline are commonly found in turns. Some preferred substitutions may be made among the following groups: (i) S and T; (ii) P and G; and (iii) A, V, L and I. Given the known genetic code, and recombinant and synthetic DNA techniques, the skilled scientist readily can construct DNAs encoding the conservative amino acid variants.

As used herein, "sequence identity" between two polypeptide sequences indicates the percentage of amino acids that are identical between the sequences. "Sequence similarity" indicates the percentage of amino acids that either are identical or that represent conservative amino acid substitutions.

Functionally Equivalent Variants

Yet another class of DNA variants within the scope of the invention may be described with reference to the product they encode. As shown below, some of the inventive DNA molecules encode a protein having a degree of homology with known proteins, or protein domains. It is expected, therefore, that they will have some or all of the requisite functional features of such molecules. These "functionally equivalent variants" products are characterized by the fact that they are functionally equivalent, with respect to biological activity, to certain known molecules.

The instant invention provides information on common structural motifs, including consensus sequences that will guide the artisan in constructing functionally equivalent variants. It will be understood that the motifs, identified for each inventive protein, may be modified within the identified consensus sequences. Thus, the invention contemplates the proteins disclosed herein that contain variability in the consensus sequences identified, and the invention further contemplates the full range of nucleic acids encoding them, and the complements of those nucleic acids.

Hybridizing Variants

DNA variants within the invention also may be described by reference to their physical properties in hybridization. One skilled in the field will recognize that DNA can be used to identify its complement and, since DNA is double stranded, its equivalent or homolog, using nucleic acid hybridization techniques. It will also be recognized that hybridization can occur with less than 100% complementarity. However, given appropriate choice of conditions, hybridization techniques can be used to differentiate among DNA sequences based on their structural relatedness to a particular probe. For guidance regarding such conditions see, for example, Sambrook *et al.*, 1989, MOLECULAR CLONING, A LABORATORY MANUAL, Cold Spring Harbor Press, N.Y.; and Ausubel *et al.*, 1989, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Green Publishing Associates and Wiley Interscience, N.Y.

Structural relatedness between two polynucleotide sequences can be expressed as a function of "stringency" of the conditions under which the two sequences will hybridize with one another. As used herein, the term "stringency" refers to the extent that the conditions disfavor hybridization. Stringent conditions strongly disfavor hybridization, and only the most structurally related molecules will hybridize to one another under such conditions. Conversely, non-stringent conditions favor hybridization of molecules displaying a lesser degree of structural relatedness. Hybridization stringency, therefore, directly correlates with the structural relationships of two nucleic acid sequences. The following relationships are useful in correlating hybridization and relatedness (where T_m is the melting temperature of a nucleic acid duplex):

a. $T_m = 69.3 + 0.41(G+C)\%$

b. The T_m of a duplex DNA decreases by 1°C with every increase of 1% in the number of mismatched base pairs.

c. $(T_m)_{\mu 2} - (T_m)_{\mu 1} = 18.5 \log_{10}\mu 2/\mu 1$
where $\mu 1$ and $\mu 2$ are the ionic strengths of two solutions.

Hybridization stringency is a function of many factors, including overall DNA concentration, ionic strength, temperature, probe size and the presence of agents which disrupt hydrogen bonding. Factors promoting hybridization include high DNA

concentrations, high ionic strengths, low temperatures, longer probe size and the absence of agents that disrupt hydrogen bonding.

Hybridization usually is done in two stages. First, in the "binding" stage, the probe is bound to the target under conditions favoring hybridization. Stringency is usually controlled at this stage by altering the temperature. For high stringency, the temperature is usually between 65°C and 70°C, unless short (<20 nt) oligonucleotide probes are used. A representative hybridization solution comprises 6X SSC, 0.5% SDS, 5X Denhardt's solution and 100μg of non-specific carrier DNA. See Ausubel *et al.*, *supra*, section 2.9, supplement 27 (1994). Of course many different, yet functionally equivalent, buffer conditions are known. Where the degree of relatedness is lower, a lower temperature may be chosen. Low stringency binding temperatures are between about 25°C and 40°C. Medium stringency is between at least about 40°C to less than about 65°C. High stringency is at least about 65°C.

Second, the excess probe is removed by washing. It is at this stage that more stringent conditions usually are applied. Hence, it is this "washing" stage that is most important in determining relatedness via hybridization. Washing solutions typically contain lower salt concentrations. One exemplary medium stringency solution contains 2X SSC and 0.1% SDS. A high stringency wash solution contains the equivalent (in ionic strength) of less than about 0.2X SSC, with a preferred stringent solution containing about 0.1X SSC. The temperatures associated with various stringencies are the same as discussed above for "binding." The washing solution also typically is replaced a number of times during washing. For example, typical high stringency washing conditions comprise washing twice for 30 minutes at 55° C. and three times for 15 minutes at 60° C.

The present invention includes nucleic acid molecules that hybridize to the inventive molecules under high stringency binding and washing conditions. More preferred molecules (from an mRNA perspective) are those that are at least 50 % of the length of any one of those depicted in below. Particularly preferred molecules are at least 75 % of the length of those molecules.

Substitutions, Insertions, Additions and Deletions

In a general sense, the preferred DNA variants of the invention are those that retain the closest relationship, as described by "sequence identity" to the inventive DNA molecules. According to another aspect of the invention, therefore, substitutions, insertions, additions and deletions of defined properties are contemplated. It will be recognized that sequence

identity between two polynucleotide sequences, as defined herein, generally is determined with reference to the protein coding region of the sequences. Thus, this definition does not at all limit the amount of DNA, such as vector DNA, that may be attached to the molecules described herein. Preferred DNA sequence variants include molecules encoding proteins sharing some or all of any relevant biological activity of the native molecule.

In creating these variants, the skilled worker will be guided by reference to the protein structure. First, insertions and deletions in any recognized functional domain, above, generally should be avoided, except as noted below in the section entitled "Proteins," where this domain is discussed in detail. Alterations in such domains usually will be limited to conservative amino acid substitutions. In addition, where insertions and deletions are desired, this may be accomplished at the N- and/or C-terminus of the protein molecule (or the corresponding coding regions of the DNA). If insertions or deletions are made within the protein, deletions of major structural features usually should be avoided. Thus, a preferred place to make insertion or deletion variants is in non-structural regions, such as linker regions between two alpha helices.

"Substitutions" generally refer to alterations in the DNA sequence which do not change its overall length, but only alter one or more nucleotide positions, substituting one for another in the common sense of the word. One class of preferred substitutions, "degenerate substitutions," are those that do not alter the encoded amino acid sequence. Some subsitutions retains 50%, 55%, 60% or 65% identity. Preferred substitutions retain at least about 70% identity, more preferably at least 70% or 75% identity, with the inventive DNAs. Some more preferred molecules have at least about 80% identity, more preferably at least 80% or 85% identity. Particularly preferred DNAs share at least about 90% identity, more preferably at least 90% or 95% identity.

"Insertions," unlike substitutions, alter the overall length of the DNA molecule, and thus sometimes the encoded protein. Insertions add extra nucleotides to the interior (not the 5' or 3' ends) of the subject DNAs. Preferred insertions are made with reference to the protein sequence encoded by the DNA. Thus, it is most preferred to provide an insertion in the DNA at a location that corresponds to an area of the encoded protein which lacks structure. For instance, it typically would not be beneficial, if the preservation of biological activity is desired, to provide an insertion within an alpha-helical region or a beta-pleated sheet. Accordingly, non-structural areas, such as those containing helix-breaking glycines

and proline residues, are most preferred sites of insertion. Other preferred sites of insertion are the splice sites, which are indicated above in the description of the inventive DNA molecules.

While the optimal size of insertions will vary depending upon the site of insertion and its effect on the overall conformation of the encoded protein, some general guides are useful. Generally, the total insertions (irrespective of their number) should not add more than about 30% (or preferably not more than 30%) to the overall size of the encoded protein. More preferably, the insertion adds less than about 10-20% (yet more preferably 10-20%) in size, with less than about 10% being most preferred. The number of insertions is limited only by the number of suitable insertions sites, and secondarily by the foregoing size preferences.

"Additions," like insertions, also add to the overall size of the DNA molecule, and usually the encoded protein. However, instead of being made within the molecule, they are made on the 5' or 3' end, usually corresponding to the N- or C- terminus of the encoded protein. Unlike deletions, additions are not very size-dependent. Indeed, additions may be of virtually any size. Preferred additions, however, do not exceed about 100% of the size of the native molecule. More preferably, they add less than about 60 to 30% to the overall size, with less than about 30% being most preferred.

"Deletions" diminish the overall size of the DNA and, therefore, also reduce the size of the protein encoded by that DNA. Deletions may be made from either end of the molecule or internal to it. Typical preferred deletions remove discrete structural features of the encoded protein. For example, some deletions will comprise the deletion of one or more exons which may define a structural feature. Preferred deletions remove less than about 30% of the size of the subject molecule. More preferred deletions remove less than about 20% and most preferred deletions remove less than about 10%.

Computer-Defined Variants and Definition of "Sequence Identity"

In general, both the DNA and protein molecules of the invention can be defined with reference to "sequence identity." As used herein, "sequence identity" refers to a comparison made between two molecules using, for example, the standard Smith-Waterman algorithm that is well known in the art.

Some molecules have at lease about 50%, 55% or 60% identity. Preferred molecules are those having at least about 65% sequence identity, more preferably at least 65% or 70% sequence identity. Other preferred molecules have at least about 80%, more preferably at

least 80% or 85%, sequence identity. Particularly preferred molecules have at least about 90% sequence identity, more preferably at least 90% sequence identity. Most preferred molecules have at least about 95%, more preferably at least 95%, sequence identity. As used herein, two nucleic acid molecules or proteins are said to "share significant sequence identity" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) identity.

"Sequence identity" is defined herein with reference the Blast 2 algorithm, which is available at the NCBI (http://www.ncbi.nlm.nih.gov/BLAST), using default parameters. References pertaining to this algorithm include: those found at http://www.ncbi.nlm.nih.gov/BLAST/blast_references.html; Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410; Gish, W. & States, D.J. (1993) "Identification of protein coding regions by database similarity search." Nature Genet. 3:266-272; Madden, T.L., Tatusov, R.L. & Zhang, J. (1996) "Applications of network BLAST server" Meth. Enzymol. 266:131-141; Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402; and Zhang, J. & Madden, T.L. (1997) "PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation." Genome Res. 7:649-656.

METHODS OF MAKING VARIANTS

It will be recognized that variants of the inventive molecules can be constructed in several different ways. For example, they may be constructed as completely synthetic DNAs. Methods of efficiently synthesizing oligonucleotides in the range of 20 to about 150 nucleotides are widely available. See Ausubel *et al.*, *supra*, section 2.11, Supplement 21 (1993). Overlapping oligonucleotides may be synthesized and assembled in a fashion first reported by Khorana *et al.*, J. Mol. Biol. 72:209-217 (1971); see also Ausubel *et al*, Section 8.2. The synthetic DNAs are designed with convenient restriction sites engineered at the 5' and 3' ends of the gene to facilitate cloning into an appropriate vector.

An alternative method of generating variants is to start with one of the inventive DNAs and then to conduct site-directed mutagenesis. See Ausubel *et al.*, *supra*, chapter 8, Supplement 37 (1997). In a typical method, a target DNA is cloned into a single-stranded

DNA bacteriophage vehicle. Single-stranded DNA is isolated and hybridized with a oligonucleotide containing the desired nucleotide alteration(s). The complementary strand is synthesized and the double stranded phage is introduced into a host. Some of the resulting progeny will contain the desired mutant, which can be confirmed using DNA sequencing. In addition, various methods are available that increase the probability that the progeny phage will be the desired mutant. These methods are well known to those in the field and kits are commercially available for generating such mutants.

ISOLATING HOMOLOGS

Methods

By using the sequences disclosed herein as probes or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs. "Homologs" are essentially naturally-occurring variants and include allelic, species-specific and tissue-specific variants.

Region-specific primers or probes derived from the nucleotide sequence(s) provided can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog using known methods (Innis *et al., PCR Protocols,* Academic Press, San Diego, CA (1990)). Such an application is useful in diagnostic methods, as described in more detail below, as well as in preparing full-length DNAs from various sources. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. As a general guide, the formula $3(G+C) + 2(A+T) = {}^{\circ}C$, is useful.

When using primers derived from the inventive sequences, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, annealing at 50-60°C), only sequences with greater than 75% sequence identity to the primer will be amplified. By employing lower stringency conditions (e.g., annealing at 35-37°C), sequences which have greater than 40-50% sequence identity to the primer also will be amplified.

The PCR product may be subcloned and sequenced to confirm that it indeed displays the expected sequence identity. The PCR fragment may then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment may be labeled

and used to screen a bacteriophage cDNA library. Alternatively, the labeled fragment may be used to screen a genomic library.

PCR technology may also be utilized to isolate full length cDNA sequences. For example, RNA may be isolated, following standard procedures, from an appropriate cellular or tissue source. A reverse transcription reaction may be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" with guanines using a standard terminal transferase reaction, the hybrid may be digested with RNAase H, and second strand synthesis may then be primed with a poly-C primer. Thus, cDNA sequences upstream of the amplified fragment may easily be isolated. For a review of cloning strategies which may be used, see e.g., Sambrook et al., 1989, *supra.*

When using DNA probes derived from the inventive sequences for colony/plaque hybridization, one skilled in the art will recognize that by employing medium to high stringency conditions (e.g., hybridizing at 50-65°C in 5X SSPC and 50% formamide, and washing at 50-65°C in 0.5X SSPC), sequences having regions with greater than 90% sequence identity to the probe can be obtained, and that by employing lower stringency conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in SSPC), sequences having regions with greater than 35-45% sequence identity to the probe will be obtained.

Suitably, genomic or cDNA libraries can be constructed and screened in accord with the previous paragraph. The libraries should be derived from a tissue or organism that is known to express the gene of interest, or that is suspected of expressing the gene. The clone containing the homolog may then be purified through methods routinely practiced in the art, and subjected to sequence analysis.

Additionally, an expression library can be constructed utilizing DNA isolated from or cDNA synthesized from a tissue or organism that is known to express the gene of interest, or that is suspected of expressing the gene. In this manner, clones may be induced and screened using standard antibody screening techniques in conjunction with antibodies raised against the normal gene product, as described herein. (For screening techniques, see, for example, Harlow, E. and Lane, eds., 1988, ANTIBODIES: A LABORATORY MANUAL, Cold Spring Harbor Press, Cold Spring Harbor Press.)

Human Homologs

Any organism or tissue can be used as the source for homologs of the present invention so long as the organism or tissue naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs is human.

PROTEINS OF THE INVENTION

One class of proteins included within the invention is encoded by the inventive DNA molecules presented. Other proteins according to the invention are those encoded by the DNA variants described above. As noted, these variants are designed with the encoded proteins in mind.

A preferred class of protein fragments includes those fragments which retain any biological activity. These molecules share functional features common the family of proteins, although these characteristics may vary in degree.

According to one aspect of the invention fragments of the inventive proteins are contemplated. Some preferred fragments are those which are capable of eliciting an immune response. Generally these "antigenic" fragments will be from about five amino acids in length to about fifty amino acids in length. Some preferred antigenic fragments are from five to about twenty amino acids long. "Antigenic" response may refer to a T cell response, a B cell response or a response by cells of the macrophage/monocyte lineages. In most cases, however, it will refer to the immune response involved in the generation of antibodies. In other words, the relevant immune response is that of helper T cells and/or B cells. These preferred molecules comprise one or more T cell and /or B cell epitopes.

ANTIBODIES OF THE INVENTION

Antibodies raised against the proteins and protein fragments of the invention also are contemplated by the invention. Described below are antibody products and methods for producing antibodies capable of specifically recognizing one or more epitopes of the presently described proteins and their derivatives.

Antibodies include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies including single chain Fv (scFv) fragments, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, epitope-binding fragments, and humanized forms of any of the above.

As known to one in the art, these antibodies may be used, for example, in the detection of a target protein in a biological sample. They also may be utilized as part of treatment methods, and/or may be used as part of diagnostic techniques whereby patients may be tested for abnormal levels or for the presence of abnormal forms of the such proteins.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A.M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth et al., *J. Immunol. Methods* 35:1-21 (1980); Kohler and Milstein, *Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* 4:72 (1983); Cole *et al.*, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985), pp. 77-96). Antibodies may also be generated by the known techniques of phage display and *in vitro* immunization.

Polyclonal Antibodies

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as an inventive protein or an antigenic derivative thereof.

Polyclonal antiserum, containing antibodies to heterogeneous epitopes of a single protein, can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified, as known in the art, to enhance immunogenicity. Immunization methods include subcutaneous or intraperitoneal injection of the polypeptide.

Effective polyclonal antibody production is affected by many factors related both to the antigen and to the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and/or adjuvant. In addition, host animal response may vary with site of inoculation. Both inadequate or excessive doses of antigen may result in low titer antisera. In general, however, small doses (high ng to low μg levels) of antigen administered at multiple intradermal sites appears to be most reliable. Host animals may include but are not limited to rabbits, mice, chickens and rats, to name but a few. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al., J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

The protein immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin β-galactosidase) or through the inclusion of an adjuvant during immunization. Adjuvants include Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum.*

Booster injections can be given at regular intervals, with at least one usually being required for optimal antibody production. The antiserum may be harvested when the antibody titer begins to fall. Titer may be determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen. See, for example, Ouchterlony *et al.*, Chap. 19 in: *Handbook of Experimental Immunology,* Wier, ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μM). The antiserum may be purified by affinity chromatography using the immobilized immunogen carried on a solid support. Such methods of affinity chromatography are well known in the art.

Affinity of the antisera for the antigen may be determined by preparing competitive binding curves, as described, for example, by Fisher, Chap. 42 in: *Manual of Clinical Immunology,* second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D.C. (1980).

In addition to using protein an the immunogen, DNA molecules may be used directly. In this manner, a DNA encoding the protein immunogen is administered. Boosting and harvesting is done in a manner analogous to that detailed above. Yet another method of producing antibodies entails immunizing chickens and harvesting the antibodies from their eggs.

Monoclonal Antibodies

Monoclonal antibodies (MAbs), are homogeneous populations of antibodies to a particular antigen. They may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture or *in vivo*. MAbs may be produced

by making hybridomas which are immortalized cells capable of secreting a specific monoclonal antibody.

Monoclonal antibodies to any of the proteins, peptides and epitopes thereof described herein can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature* 256:495-497 (1975) (and U.S. Patent No. 4,376,110) or modifications of the methods thereof, such as the human B-cell hybridoma technique (Kosbor *et al.*, 1983, *Immunology Today* 4:72; Cole *et al.*, 1983, *Proc. Natl. Acad. Sci.* USA 80: 2026-2030), and the EBV-hybridoma technique (Cole *et al.*, 1985, MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In one method a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen are isolated.

The spleen cells are fused, typically using polyethylene glycol, with mouse myeloma cells, such as SP2/0-Ag14 myeloma cells. The excess, unfused cells are destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted, and aliquots are plated to microliter plates where growth is continued.

Antibody-producing clones (hybridomas) are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures. These include ELISA, as originally described by Engvall, *Meth. Enzymol.* 70:419 (1980), western blot analysis, radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175:109-124 (1988)) and modified methods thereof.

Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* BASIC METHODS IN MOLECULAR BIOLOGY, Elsevier, New York. Section 21-2 (1989). The hybridoma clones may be cultivated *in vitro* or *in vivo*, for instance as ascites. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production. Alternatively, hybridoma culture in hollow fiber bioreactors provides a continuous high yield source of monoclonal antibodies.

The antibody class and subclass may be determined using procedures known in the art (Campbell, A.M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

MAbs may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. Methods of purifying monoclonal antibodies are well known in the art.

Antibody Derivatives and Fragments

Fragments or derivatives of antibodies include any portion of the antibody which is capable of binding the target antigen, or a specific portion thereof. Antibody derivatives include poly-specific (*e.g.*, bi-specific) antibodies, which contain binding sites specific for two or more different epitopes. These epitopes may be from the same or different inventive molecules or one or more epitope may be from a molecule not specifically disclosed here.

Antibody fragments specifically include $F(ab')_2$, Fab, Fab' and Fv fragments. These can be generated from any class of antibody, but typically are made from IgG or IgM. They may be made by conventional recombinant DNA techniques or, using the classical method, by proteolytic digestion with papain or pepsin. See CURRENT PROTOCOLS IN IMMUNOLOGY, chapter 2, Coligan *et al.*, eds., (John Wiley & Sons 1991-92).

$F(ab')_2$ fragments are typically about 110 kDa (IgG) or about 150 kDa (IgM) and contain two antigen-binding regions, joined at the hinge by disulfide bond(s). Virtually all, if not all, of the Fc is absent in these fragments. Fab' fragments are typically about 55 kDa (IgG) or about 75 kDa (IgM) and can be formed, for example, by reducing the disulfide bond(s) of an $F(ab')_2$ fragment. The resulting free sulfhydryl group(s) may be used to conveniently conjugate Fab' fragments to other molecules, such as detection reagents (*e.g.*, enzymes).

Fab fragments are monovalent and usually are about 50 kDa (from any source). Fab fragments include the light (L) and heavy (H) chain, variable (V_L and V_H, respectively) and constant (C_L C_H, respectively) regions of the antigen-binding portion of the antibody. The H and L portions are linked by an intramolecular disulfide bridge.

Fv fragments are typically about 25 kDa (regardless of source) and contain the variable regions of both the light and heavy chains (V_L and V_H, respectively). Usually, the V_L and V_H chains are held together only by non-covalent interacts and, thus, they readily dissociate. They do, however, have the advantage of small size and they retain the same binding properties of the larger Fab fragments. Accordingly, methods have been developed to crosslink the V_L and V_H chains, using, for example, glutaraldehyde (or other chemical crosslinkers), intermolecular disulfide bonds (by incorporation of cysteines) and peptide linkers. The resulting Fv is now a single chain (*i.e.*, SCFv).

Other antibody derivatives include single chain antibodies (U.S. Patent 4,946,778; Bird, Science 242:423-426 (1988); Huston *et al.*, Proc. Natl. Acad. Sci. USA 85:5879-5883 (1988); and Ward *et al.*, Nature 334:544-546 (1989)). Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain FV (SCFv).

One preferred method involves the generation of scFvs by recombinant methods, which allows the generation of Fvs with new specificities by mixing and matching variable chains from different antibody sources. In a typical method, a recombinant vector would be provided which comprises the appropriate regulatory elements driving expression of a cassette region. The cassette region would contain a DNA encoding a peptide linker, with convenient sites at both the 5' and 3' ends of the linker for generating fusion proteins. The DNA encoding a variable region(s) of interest may be cloned in the vector to form fusion proteins with the linker, thus generating an scFv.

In an exemplary alternative approach, DNAs encoding two Fvs may be ligated to the DNA encoding the linker, and the resulting tripartite fusion may be ligated directly into a conventional expression vector. The scFv DNAs generated any of these methods may be expressed in prokaryotic or eukaryotic cells, depending on the vector chosen.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab)_2$ fragments. Alternatively, Fab expression libraries may be constructed (Huse et al., 1989, *Science*, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Derivatives also include "chimeric antibodies" (Morrison *et al.*, *Proc. Natl. Acad. Sci.*, 81:6851-6855 (1984); Neuberger *et al.*, *Nature*, 312:604-608 (1984); Takeda *et al.*, *Nature*, 314:452-454 (1985)). These chimeras are made by splicing the DNA encoding a mouse antibody molecule of appropriate specificity with, for instance, DNA encoding a human antibody molecule of appropriate specificity. Thus, a chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. These are also known sometimes as "humanized" antibodies and they offer the added

advantage of at least partial shielding from the human immune system. They are, therefore, particularly useful in therapeutic *in vivo* applications.

Labeled Antibodies

The present invention further provides the above-described antibodies in detectably labeled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.) fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well-known in the art, for example see (Sternberger *et al., J. Histochem. Cytochem.* 18:315 (1970); Bayer *et al., Meth. Enzym.* 62:308 (1979); Engval *et al., Immunol.* 109:129 (1972); Goding, *J. Immunol. Meth.* 13:215 (1976)). The labeled antibodies of the present invention can be used for *in vitro, in vivo*, and *in situ* diagnostic assays.

Immobilized Antibodies

The foregoing antibodies also may be immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir *et al.*, "*Handbook of Experimental Immunology*" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby *et al., Meth. Enzym.* 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro, in vivo*, and *in situ* assays as well as for immunoaffinity purification of the proteins of the present invention.

THERAPEUTIC AND DIAGNOSTIC COMPOSITIONS

The proteins, antibodies and polynucleotides of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, e.g., human serum albumin, are described, for example, in *Remington's Pharmaceutical Sciences* (16th ed., Osol, A., Ed., Mack, Easton PA (1980)). In order to form a pharmaceutically acceptable composition suitable for effective administration,

such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvate may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral or rectal administration.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (*e.g.*, pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (*e.g.*, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (*e.g.*, magnesium stearate, talc or silica); disintegrants (*e.g.*, potato starch or sodium starch glycolate); or wetting agents (*e.g.*, sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they maybe presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound. For buccal administration the composition may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.* gelatin for

use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

RECOMBINANT CONSTRUCTS AND EXPRESSION

The present invention further provides recombinant DNA constructs comprising one or more of the nucleotide sequences of the present invention. The recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a DNA or DNA fragment, typically bearing an open reading frame, is inserted, in either orientation.

The gene products encoded by the subject DNAs may be produced by recombinant DNA technology using techniques well known in the art. See, for example, the techniques described in Sambrook et al., 1989, *supra*, and Ausubel et al., 1989, *supra*. Alternatively, the DNA sequences may be chemically synthesized using, for example, synthesizers. See, for

example, the techniques described in OLIGONUCLEOTIDE SYNTHESIS, 1984, Gait, ed., IRL Press, Oxford, which is incorporated by reference herein in its entirety. They may be assembled from fragments and short oligonucleotide linkers, or from a series of oligonucleotides. The are preferably made by RT-PCR methods. The resulting synthetic gene is capable of being expressed in a recombinant vector.

In some cases the recombinant constructs will be expression vectors, which are capable of expressing the RNA and/or protein products of the encoded DNA(s). Thus, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the open reading frame (ORF). The vector may further comprise a selectable marker sequence.

Specific initiation signals may also be required for efficient translation of inserted target gene coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where a target DNA includes its own initiation codon and adjacent sequences is inserted into the appropriate expression vector, no additional translation control signals may be needed. However, in cases where only a portion of an ORF is used, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire target. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bittner *et al.*, *Methods in Enzymol.* 153:516-544 (1987)). Some appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, *et al., in Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U.S. Patent No. 5,082,767.

The present invention further provides host cells containing at least one of the DNAs of the present invention. The host cell can be virtually any cell for which expression vectors are available. It may be, for example, a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic

cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis *et al.*, *Basic Methods in Molecular Biology* (1986)).

A wide variety of expression systems are available, such as: yeast (*e.g. Saccharomyces, Pichia*) transformed with recombinant yeast expression vectors containing the target DNA; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing the target DNA sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.* Ti plasmid) containing target DNA coding sequences; or mammalian cell systems (*e.g.* COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter).

Depending on the system chosen, the resulting product may differ. For example, proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

Vectors

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting selection of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequence, and in one aspect of the invention, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal or C-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product.

Bacterial Expression

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, if desirable, to provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli, Bacillus subtilis, Salmonella typhimurium* and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may, also be employed as a matter of choice.

Bacterial vectors may be, for example, bacteriophage-, plasmid- or cosmid-based. These vectors can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids typically containing elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, GEM 1 (Promega Biotec, Madison, WI, USA), pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pKK232-8, pDR540, and pRIT5 (Pharmacia).

These "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Bacterial promoters include lac, T3, T7, lambda P_R or P_L, trp, and ara.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is derepressed/induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the protein being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide libraries, for example, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther et al., 1983, *EMBO J.* 2:1791), in which the coding sequence may be ligated into the vector in frame with the *lac Z* coding region so that a fusion protein is produced; pIN vectors (Inouye *et al.* 1985, *Nucleic Acids*

Res. 13:3101-3109; Van Heeke *et al.*, 1989, *J. Biol. Chem.* 264:5503-5509); pET vectors, Studier *et al.*, *Methods in Enzymology* 185: 60-89 (Academic Press 1990); and the like.

Moreover, pGEX vectors may be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and easily can be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene protein can be released from the GST moiety.

In a one embodiment, full length cDNA sequences are appended with in-frame*BamHI* sites at the amino terminus and *EcoRI* sites at the carboxyl terminus using standard PCR methodologies (Innis et al., 1990, *supra*) and ligated into the pGEX-2TK vector (Pharmacia, Uppsala, Sweden). The resulting cDNA construct contains a kinase recognition site at the amino terminus for radioactive labeling and glutathione S-transferase sequences at the carboxyl terminus for affinity purification (Nilsson, *et al.* 1985, *EMBO J.* 4: 1075; Zabeau and Stanley, 1982, *EMBO J.* 1:1217.

Eukaryotic Expression

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell 23*:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Mammalian promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Exemplary mammalian vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). Selectable markers include CAT (chloramphenicol transferase).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the coding sequence of interest

may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a target protein in infected hosts. (*E.g.*, See Logan *et al.*, 1984, *Proc. Natl. Acad. Sci.* USA 81:3655-3659).

In one embodiment, cDNA sequences encoding the full-length open reading frames are ligated into pCMVß replacing the ß-galactosidase gene such that cDNA expression is driven by the CMV promoter (Alam, 1990, *Anal. Biochem.* 188: 245-254; MacGregor *et al.*, 1989, *Nucl. Acids Res.* 17: 2365; Norton *et al.* 1985, *Mol. Cell. Biol.* 5: 281).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins.

Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

For long-term, high-yield production of recombinant proteins in eukaryotic cells, stable expression is preferred. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, *etc.*), and a selectable marker.

Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the target protein. Such engineered cell lines may be

particularly useful in screening and evaluation of compounds that affect the endogenous activity of the protein.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, *et al.*, *Cell* 11:223 (1977)), hypoxanthine-guanine phosphoribosyltransferase(Szybalska *et al.*, *Proc. Natl. Acad. Sci. USA* 48:2026 (1962)), and adenine phosphoribosyltransferase(Lowy, *et al.*, *Cell* 22:817 (1980)) genes can be employed in tk⁻, hgprt⁻ or aprt⁻ cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler, *et al.*, *Proc. Natl. Acad, Sci. USA* 77:3567 (1980)); O'Hare, *et al.*, 1981, *Proc. Natl. Acad. Sci. USA* 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2072 (1981)); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, *et al.*, 1981, *J. Mol. Biol.* 150:1); and hydro, which confers resistance to hygromycin (Santerre, *et al.*, 1984, *Gene* 30:147) genes.

An alternative fusion protein system allows for the ready purificationof non-denatured fusion proteins expressed in human cell lines (Janknecht, *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 8972-8976 (1991)). In this system, the gene of interest is subcloned into a vaccinia-based plasmid such that the gene's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni^{2+} nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containingbuffers.

In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The target coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of a target gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed. (*E.g.*, see Smith et al., 1983, *J. Virol.* 46: 584; Smith, U.S. Patent No. 4,215,051).

While the present proteins can be expressed in recombinant systems, as described above, cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

Purification of Recombinant Proteins

Recombinant proteins produced may be isolated by host cell lysis. This may be followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, like lysozyme and chelators.

If inclusion bodies are formed in bacterial systems, they may be extracted from cell pellets using, for example, detergents, reducing agents, salts, urea, guanidinium chloride and extremes of pH (*e.g.* <4 or >10). If denaturation occurs, protein refolding steps (*e.g.*, dialysis) can be used, as necessary, in completing configuration of the mature protein. If disulfide bridges are present in the native protein, they may be reoxidized using known methods.

By way of specific non-limiting example, the recombinant bacterial cells, for example *E. coli*, are grown in any of a number of suitable media, for example LB, and the expression of the recombinant protein induced by adding IPTG (*e.g.*, *lac* operator-promoter) to the media or switching incubation to a higher temperature (*e.g.*, λ cI^{857}). After culturing the bacteria for a further period of between 2 and 24 hours, the cells are collected by centrifugation and washed to remove residual media. The bacterial cells are then lysed, for example, by disruption in a cell homogenizer and centrifuged to separate the cell membranes from the soluble cell components. If the protein aggregates into inclusion bodies, this centrifugation can be performed under conditions whereby the dense inclusion bodies are selectively enriched by incorporation of sugars such as sucrose into the buffer and centrifugation at a selective speed. The inclusion bodies can then be washed in any of several solutions to remove some of the contaminating host proteins, then solubilized in solutions containing high concentrations of urea (*e.g.* 8M) or chaotropic agents such as guanidinium hydrochloride in the presence of reducing agents such as ß-mercaptoethanol or DTT (dithiothreitol).

At this stage it may be advantageous to incubate the protein for several hours under conditions suitable for the protein to undergo a refolding process into a conformation which

more closely resembles that of the native protein. Such conditions generally include low protein concentrations less than 500 μg/ml), low levels of reducing agent, concentrations of urea less than 2 M and often the presence of reagents such as a mixture of reduced and oxidized glutathione which facilitate the interchange of disulphide bonds within the protein molecule. The refolding process can be monitored, for example, by SDS-PAGE or with antibodies which are specific for the native molecule. Following refolding, the protein can then be purified further and separated from the refolding mixture by chromatography on any of several supports including ion exchange resins, gel permeation resins or on a variety of affinity columns.

Labeling Proteins

When used as a component in assay systems such as those described, below, the target protein may be labeled, either directly or indirectly, to facilitate detection of the present *res*-like molecules either *in vitro* or *in vivo*. Any of a variety of suitable labeling systems may be used including but not limited to radioisotopes such as ^{125}I; enzyme labeling systems that generate a detectable colorimetric signal or light when exposed to substrate; and fluorescent labels.

Where recombinant DNA technology is used for protein production the, it may be advantageous to engineer fusion proteins that can facilitate labeling, immobilization and/or detection. These fusion proteins may, for example, add amino acids which facilitate further chemical modification. They also may add a functional moiety, such as an enzyme, which directly facilitates detection.

TRANSGENIC ANIMALS

The invention further contemplates animal models for studying the function of the present molecules and for overproducing the protein products. The disclosed DNA sequences may be used in conjunction with techniques for producing transgenic animals that are well known to those of skill in the art.

To prepare transgenic animals, target gene sequences may for example be introduced into, and overexpressed in, the genome of the animal of interest, or, if endogenous target gene sequences are present, they may either be overexpressed or, alternatively, be disrupted in order to underexpress or inactivate target gene expression, such as described for the disruption of apoE in mice (Plum *et al.*, *Cell* 71: 343-353 (1992)).

In order to overexpress a target gene sequence, the coding portion of the target gene sequence may be ligated to a regulatory sequence which is capable of driving gene expression in the animal and cell type of interest. Such regulatory regions will be well known to those of skill in the art, and may be utilized in the absence of undue experimentation.

For underexpression of an endogenous target gene sequence, such a sequence may be isolated and engineered such that when reintroduced into the genome of the animal of interest, the endogenous target gene alleles will be inactivated. Preferably, the engineered target gene sequence is introduced via gene targeting such that the endogenous target sequence is disrupted upon integration of the engineered target gene sequence into the animal's genome.

Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate cardiovascular disease animal models. Goats, cows and sheep are particularly preferred for producing protein *in vivo*.

Any technique known in the art may be used to introduce a target gene transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe *et al.*, U.S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82:6148-6152 (1985)); gene targeting in embryonic stem cells (Thompson *et al.*, *Cell* 56:313-321 (1989)); electroporation of embryos (Lo, *Mol. Cell. Biol.* 3:1803-1814 (1983)); and sperm-mediated gene transfer (Lavitrano *et al.*, *Cell* 57:717-723 (1989)); *etc.* For a review of such techniques, see Gordon, Transgenic Animals, *Intl. Rev. Cytol.* 115:171-229 (1989).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals. The transgene may be integrated as a single transgene or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching

of Lasko et al. (Lasko *et al.*, *Proc. Natl. Acad. Sci.* USA 89:3232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the target gene be integrated into the chromosomal site of the endogenous target gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous target gene of interest are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous target gene.

The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene of interest in only that cell type, by following, for example, the teaching of Gu *et al. Science* 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant target gene and protein may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include but are not limited to Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and RT-PCR. Samples of target gene-expressing tissue, may also be evaluated immunocytochemically using antibodies specific for the target gene transgene gene product of interest.

The transgenic animals that express target gene mRNA or target gene transgene peptide (detected immunocytochemically, using antibodies directed against the target gene product's epitopes) at easily detectable levels should then be further evaluated to identify those animals which display characteristic increased susceptibility to carcinogenesis. Additionally, specific cell types within the transgenic animals may be analyzed and assayed *in vitro* for cellular phenotypes characteristic of mutant phenotype.

Once target gene transgenic founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include but are not limited to: outbreeding of founder animals with more

than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound target gene transgenics that express the target gene transgene of interest at higher levels because of the effects of additive expression of each target gene transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order both to augment expression and eliminate the possible need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; breeding animals to different inbred genetic backgrounds so as to examine effects of modifying alleles on expression of the target gene transgene and the possible development of carcinogenesis. One such approach is to cross the target gene transgenic founder animals with a wild type strain to produce an F1 generation that exhibits increased susceptibility to carcinogenesis. The F1 generation may then be inbred in order to develop a homozygous line, if it is found that homozygous target gene transgenic animals are viable.

Methods of generating "knockout" mice using homologous recombination in embryonic stem cells are well known in the art. Suitable methods are described, for example, in Mansour *et al.*, *Nature*, 336:348 (1988); Zijlstra *et al.*, *Nature*, 342:435 (1989) and 344:742 (1990); and Hasty *et al.*, *Nature*, 350:243 (1991). This genomic DNA can be obtained by conventional methods using the cDNA sequence as a probe in a commercially-available genomic DNA library.

Briefly, a genomic fragment is cleaved with a restriction endonuclease and a heterologous cassette containing a neomycin-resistance gene is inserted at the cleavage site. A suitable cassette is the GTI-II *neo* cassette described by Lufkin *et al.*, *Cell* 66:1105 (1991). The modified genomic fragment is cloned into a suitable targeting vector that is introduced into murine embryonic stem cells by electroporation. Cells that have undergone homologous recombination (and hence disruption of the gene) are selected by resistance to G418, and used to generate chimeric mice using well known methods. See Lufkin *et al.*, *supra*. Traditional breeding methods then can be used to generate mice that are homozygous for the disrupted gene.

The phenotype of mice that are homozygous for the mutation then can be studied to provide insights into the role of the protein in, for example, carcinogenesis. These mice also can be used as models for developing new treatments for cancers. If this mutation is lethal in

homozygous mice (for example during embryogenesis) heterozygous mice, which express only half the amount of the protein can also be studied.

GENE THERAPY APPLICATIONS

When mutations in the inventive protein, or in the elements controlling expression of that protein, are found to be associated with a malignant phenotype, control of cellular proliferation can be restored by gene therapy methods. For example, overexpression of the protein can be counteracted by concurrent expression of an antisense molecule that binds to and inhibits expression of the mRNA encoding the protein. Alternatively, overexpression can be inhibited in an analogous manner using a ribozyme that cleaves the mRNA. In another embodiment, where expression of a mutated protein induces the malignant phenotype, concomitant expression of the non-mutated molecule via introduction of an exogenous gene may be used. Methods of using antisense and ribozyme technology to control gene expression, or of gene therapy methods for expression of an exogenous gene in this manner are well known in the art.

Each of these methods requires a system for introducing a vector into the cells containing the mutated gene. The vector encodes either an antisense or ribozyme transcript of the inventive protein. The construction of a suitable vector can be achieved by any of the methods well-known in the art for the insertion of exogenous DNA into a vector. *See, e.g.*, Sambrook *et al.*, *Molecular Cloning* (Cold Spring Harbor Press 2d ed. 1989), which is incorporated herein by reference. In addition, the prior art teaches various methods of introducing exogenous genes into cells *in vivo*. *See* Rosenberg *et al.*, *Science 242*:1575-1578 (1988) and Wolff *et al.*, *PNAS* 86:9011-9014 (1989), which are incorporated herein by reference. The routes of delivery include systemic administration and administration *in situ*. Well-known techniques include systemic administration with cationic liposomes, and administration *in situ* with viral vectors. Any one of the gene delivery methodologies described in the prior art is suitable for the introduction of a recombinant vector containing an inventive gene according to the invention into a MTX-resistant, transport-deficient cancer cell. A listing of present-day vectors suitable for the purpose of this invention is set forth in Hodgson, *Bio/Technology 13*: 222 (1995), which is incorporated by reference.

For example, liposome-mediated gene transfer is a suitable method for the introduction of a recombinant vector containing an inventive gene according to the invention

into a MTX-resistant, transport-deficient cancer cell. The use of a cationic liposome, such as DC-Chol/DOPE liposome, has been widely documented as an appropriate vehicle to deliver DNA to a wide range of tissues through intravenous injection of DNA/cationic liposome complexes. *See* Caplen *et al.*, *Nature Med.* 1:39-46 (1995) and Zhu *et al.*, *Science 261*:209-211 (1993), which are herein incorporated by reference. Liposomes transfer genes to the target cells by fusing with the plasma membrane. The entry process is relatively efficient, but once inside the cell, the liposome-DNA complex has no inherent mechanism to deliver the DNA to the nucleus. As such, the most of the lipid and DNA gets shunted to cytoplasmic waste systems and destroyed. The obvious advantage of liposomes as a gene therapy vector is that liposomes contain no proteins, which thus minimizes the potential of host immune responses.

As another example, viral vector-mediated gene transfer is also a suitable method for the introduction of the vector into a target cell. Appropriate viral vectors include adenovirus vectors and adeno-associated virus vectors, retrovirus vectors and herpesvirus vectors.

Adenoviruses are linear, double stranded DNA viruses complexed with core proteins and surrounded by capsid proteins. The common serotypes 2 and 5, which are not associated with any human malignancies, are typically the base vectors. By deleting parts of the virus genome and inserting the desired gene under the control of a constitutive viral promoter, the virus becomes a replication deficient vector capable of transferring the exogenous DNA to differentiated, non-proliferating cells. To enter cells, the adenovirus fibre interacts with specific receptors on the cell surface, and the adenovirus surface proteins interact with the cell surface integrins. The virus penton-cell integrin interaction provides the signal that brings the exogenous gene-containing virus into a cytoplasmic endosome. The adenovirus breaks out of the endosome and moves to the nucleus, the viral capsid falls apart, and the exogenous DNA enters the cell nucleus where it functions, in an epichromosomal fashion, to express the exogenous gene. Detailed discussions of the use of adenoviral vectors for gene therapy can be found in Berkner, *Biotechniques* 6:616-629 (1988) and Trapnell, *Advanced Drug Delivery Rev. 12*:185-199 (1993), which are herein incorporated by reference. Adenovirus-derived vectors, particularly non-replicative adenovirus vectors, are characterized by their ability to accommodate exogenous DNA of 7.5 kB, relative stability, wide host range, low pathogenicity in man, and high titers (10^4 to 10^5 plaque forming units per cell). *See* Stratford-Perricaudet *et al.*, *PNAS 89*:2581 (1992).

Adeno-associated virus (AAV) vectors also can be used for the present invention. AAV is a linear single-stranded DNA parvovirus that is endogenous to many mammalian species. AAV has a broad host range despite the limitation that AAV is a defective parvovirus which is dependent totally on either adenovirus or herpesvirus for its reproduction *in vivo*. The use of AAV as a vector for the introduction into target cells of exogenous DNA is well-known in the art. *See, e.g.*, Lebkowski *et al.*, *Mole. & Cell. Biol.* 8:3988 (1988), which is incorporated herein by reference. In these vectors, the capsid gene of AAV is replaced by a desired DNA fragment, and transcomplementation of the deleted capsid function is used to create a recombinant virus stock. Upon infection the recombinant virus uncoats in the nucleus and integrates into the host genome.

Another suitable virus-based gene delivery mechanism is retroviral vector-mediated gene transfer. In general, retroviral vectors are well-known in the art. *See* Breakfield *et al.*, *Mole. Neuro. Biol.* *1*:339 (1987) and Shih *et al.*, in *Vaccines* 85: 177 (Cold Spring Harbor Press 1985). A variety of retroviral vectors and retroviral vector-producing cell lines can be used for the present invention. Appropriate retroviral vectors include Moloney Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus. These vectors include replication-competent and replication-defective retroviral vectors. In addition, amphotropic and xenotropic retroviral vectors can be used. In carrying out the invention, retroviral vectors can be introduced to a tumor directly or in the form of free retroviral vector producing-cell lines. Suitable producer cells include fibroblasts, neurons, glial cells, keratinocytes, hepatocytes, connective tissue cells, ependymal cells, chromaffin cells. *See* Wolff *et al.*, *PNAS 84*:3344 (1989).

Retroviral vectors generally are constructed such that the majority of its structural genes are deleted or replaced by exogenous DNA of interest, and such that the likelihood is reduced that viral proteins will be expressed. *See* Bender *et al.*, *J. Virol.* 61:1639 (1987) and Armento *et al.*, *J. Virol. 61*:1647 (1987), which are herein incorporated by reference. To facilitate expression of the antisense or ribozyme molecule, of the inventive protein, a retroviral vector employed in the present invention must integrate into the genome of the host cell genome, an event which occurs only in mitotically active cells. The necessity for host cell replication effectively limits retroviral gene expression to tumor cells, which are highly

replicative, and to a few normal tissues. The normal tissue cells theoretically most likely to be transduced by a retroviral vector, therefore, are the endothelial cells that line the blood vessels that supply blood to the tumor. In addition, it is also possible that a retroviral vector would integrate into white blood cells both in the tumor or in the blood circulating through the tumor.

The spread of retroviral vector to normal tissues, however, is limited. The local administration to a tumor of a retroviral vector or retroviral vector producing cells will restrict vector propagation to the local region of the tumor, minimizing transduction, integration, expression and subsequent cytotoxic effect on surrounding cells that are mitotically active.

Both replicatively deficient and replicatively competent retroviral vectors can be used in the invention, subject to their respective advantages and disadvantages. For instance, for tumors that have spread regionally, such as lung cancers, the direct injection of cell lines that produce replication-deficient vectors may not deliver the vector to a large enough area to completely eradicate the tumor, since the vector will be released only form the original producer cells and their progeny, and diffusion is limited. Similar constraints apply to the application of replication deficient vectors to tumors that grow slowly, such as human breast cancers which typically have doubling times of 30 days versus the 24 hours common among human gliomas. The much shortened survival-time of the producer cells, probably no more than 7-14 days in the absence of immunosuppression, limits to only a portion of their replicative cycle the exposure of the tumor cells to the retroviral vector.

The use of replication-defective retroviruses for treating tumors requires producer cells and is limited because each replication-defective retrovirus particle can enter only a single cell and cannot productively infect others thereafter. Because these replication-defective retroviruses cannot spread to other tumor cells, they would be unable to completely penetrate a deep, multilayered tumor *in vivo*. *See* Markert *et al.*, *Neurosurg. 77*: 590 (1992). The injection of replication-competent retroviral vector particles or a cell line that produces a replication-competent retroviral vector virus may prove to be a more effective therapeutic because a replication competent retroviral vector will establish a productive infection that will transduce cells as long as it persists. Moreover, replicatively competent retroviral vectors may follow the tumor as it metastasizes, carried along and propagated by transduced tumor cells. The risks for complications are greater, with replicatively competent vectors, however.

Such vectors may pose a greater risk then replicatively deficient vectors of transducing normal tissues, for instance. The risks of undesired vector propagation for each type of cancer and affected body area can be weighed against the advantages in the situation of replicatively competent verses replicatively deficient retroviral vector to determine an optimum treatment.

Both amphotropic and xenotropic retroviral vectors may be used in the invention. Amphotropic viruses have a very broad host range that includes most or all mammalian cells, as is well known to the art. Xenotropic viruses can infect all mammalian cell cells except mouse cells. Thus, amphotropic and xenotropic retroviruses from many species, including cows, sheep, pigs, dogs, cats, rats, and mice, *inter alia* can be used to provide retroviral vectors in accordance with the invention, provided the vectors can transfer genes into proliferating human cells *in vivo*.

Clinical trials employing retroviral vector therapy treatment of cancer have been approved in the United States. *See* Culver, *Clin. Chem. 40*: 510 (1994). Retroviral vector-containing cells have been implanted into brain tumors growing in human patients. *See* Oldfield *et al., Hum. Gene Ther. 4*: 39 (1993). These retroviral vectors carried the HSV-1 thymidine kinase (HSV-tk) gene into the surrounding brain tumor cells, which conferred sensitivity of the tumor cells to the antiviral drug ganciclovir. Some of the limitations of current retroviral based cancer therapy, as described by Oldfield are: (1) the low titer of virus produced, (2) virus spread is limited to the region surrounding the producer cell implant, (3) possible immune response to the producer cell line, (4) possible insertional mutagenesis and transformation of retroviral infected cells, (5) only a single treatment regimen of pro-drug, ganciclovir, is possible because the "suicide" product kills retrovirally infected cells and producer cells and (6) the bystander effect is limited to cells in direct contact with retrovirally transformed cells. *See* Bi *et al.*, *Human Gene Therapy 4*: 725 (1993).

Yet another suitable virus-based gene delivery mechanism is herpesvirus vector-mediated gene transfer. While much less is known about the use of herpesvirus vectors, replication-competent HSV-1 viral vectors have been described in the context of antitumor therapy. *See* Martuza *et al., Science 252*: 854 (1991), which is incorporated herein by reference.

DIAGNOSTIC METHODS

The present invention also contemplates, for certain molecules described below, methods for diagnosis of human disease. In particular, patients can be screened for the occurrence of cancers, or likelihood of occurrence of cancers, associated with mutations in the encoded protein. DNA from tumor tissue obtained from patients suffering from cancer can be isolated and the gene encoding the protein can be sequenced. By examining a number of patients in this manner, mutations in the gene that are associated with a malignant cellular phenotype can be identified. In addition, correlation of the nature of the observed mutations with subsequent observed clinical outcomes allows development of prognostic model for the predicted outcome in a particular patient.

Screening for mutations conveniently can be carried out at the DNA level by use of PCR, although the skilled artisan will be aware that many other well known methods are available for the screening. PCR primers can be selected that flank known mutation sites, and the PCR products can be sequenced to detect the occurrence of the mutation. Alternatively, the 3' residue of one PCR primer can be selected to be a match only for the residue found in the unmutated gene. If the gene is mutated, there will be a mismatch at the 3' end of the primer, and primer extension cannot occur, and no PCR product will be obtained. Alternatively, primer mixtures can be used where the 3' residue of one primer is any nucleotide other than the nonmutated residue. Observation of a PCR product then indicates that a mutation has occurred. Other methods of using, for example, oligonucleotide probes to screen for mutations are described, or example, in U.S. Patent No. 4,871,838, which is herein incorporated by reference in its entirety.

Alternatively, antibodies can be generated that selectively bind either mutated or non-mutated protein. The antibodies then can be used to screen tissue samples for occurrence of mutations in a manner analogous to the DNA-based methods described *supra*.

The diagnostic methods described above can be used not only for diagnosis and for prognosis of existing disease, but may also be used to predict the likelihood of the future occurrence of disease. For example, clinically healthy patients can be screened for mutations in the inventive molecule that correlate with later disease onset. Such mutations may be observed in the heterozygous state in healthy individuals. In such cases a single mutation event can effectively disable proper functioning of the gene and induce a transformed or malignant phenotype. This screening also may be carried out prenatally or neonatally.

DNA molecules according to the invention also are well suited for use in so-called "gene chip" diagnostic applications. Such applications have been developed by, *inter alia*, Synteni and Affymetrix. Briefly, all or part of the DNA molecules of the invention can be used either as a probe to screen a polynucleotide array on a "gene chip," or they may be immobilized on the chip itself and used to identify other polynucleotides via hybridization to the surface of the chip. In this manner, for example, related genes can be identified, or expression patterns of the gene in various tissues can be simultaneously studied. Such gene chips have particular application for diagnosis of disease, or in forensic analysis to detect the presence or absence of an analyte. Suitable chip technology is described for example, in Wodicka *et al.*, *Nature Biotechnology*, 15:1359 (1997) which is hereby incorporated by reference in its entirety, and references cited therein.

PROTEIN-PROTEIN INTERACTIONS

Due to their similarity to certain known proteins, it is anticipated that some of the inventive protein molecules will interact with another class of cellular proteins. This is particularly true of those molecule containing leucine zipper motifs.

Any method suitable for detecting protein-protein interactions can be employed for identifying interacting targets. Among the traditional methods which can be employed are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns. Utilizing procedures such as these allows for the identification of GAP gene products. Once identified, a GAP protein can be used, in conjunction with standard techniques, to identify its corresponding pathway gene. For example, at least a portion of the amino acid sequence of the pathway gene product can be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique (see, e.g., Creighton, 1983, PROTEINS: STRUCTURES AND MOLECULAR PRINCIPLES, W.H. Freeman & Co., N.Y., pp.34-49). The amino acid sequence obtained can be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for pathway gene sequences. Screening can be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and for screening are well-known. (See *e.g.*, Ausubel, *supra*, and PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, 1990, Innis *et al.*, eds. Academic Press, Inc., New York).

Additionally, methods can be employed which result in the simultaneous identification of interacting target genes. One method which detects protein interactions *in vivo*, the two-hybrid system, is described in detail for illustration purposes only and not by way of limitation. One version of this system has been described (Chien *et al.*, *Proc. Natl. Acad. Sci. USA*, 88: 9578-9582 (1991)) and is commercially available from Clontech (Palo Alto, CA).

Briefly, utilizing such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain of a transcription activator protein fused to a known protein, in this case an inventive protein, and the other contains the activator protein's activation domain fused to an unknown protein (a putative GAP, for instance) that is encoded by a cDNA which has been recombined into this plasmid as part of a cDNA library. The plasmids are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (*e.g.*, *lacZ*) whose regulatory region contains the transcription activator's binding sites. Either hybrid protein alone cannot activate transcription of the reporter gene, the DNA-binding domain hybrid cannot because it does not provide activation function, and the activation domain hybrid cannot because it cannot localize to the activator's binding sites. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

The two-hybrid system or related methodology can be used to screen activation domain libraries for proteins that interact with a known "bait" gene product. By way of example, and not by way of limitation, gene products known to be involved in TH cell subpopulation-related disorders and/or differentiation, maintenance, and/or effector function of the subpopulations can be used as the bait gene products. Total genomic or cDNA sequences are fused to the DNA encoding on activation domain. This library and a plasmid encoding a hybrid of the bait gene product fused to the DNA-binding domain are cotransformed into a yeast reporter strain, and the resulting transformants are screened for those that express the reporter gene. For example, and not by way of limitation, the bait gene can be cloned into a vector such that it is translationally fused to the DNA encoding the DNA-binding domain of the GAL4 protein. These colonies are purified and the library plasmids responsible for reporter gene expression are isolated. DNA sequencing is then used to identify the proteins encoded by the library plasmids.

The present invention, thus generally described, will be understood more readily by reference to the following examples, which are provided by way of illustration and are not intended to be limiting of the present invention.

The examples below are provided to illustrate the subject invention. These examples are provided by way of illustration and are not included for the purpose of limiting the invention.

EXAMPLES

EXAMPLE I: cDNA Library Construction

cDNA library plates and clones originated from five cDNA libraries that were constructed by directional cloning. These are available through the Resource Center (http://www.rzpd.de) of the German Genome Project. In particular, the hfbr2 (human fetal brain; RZPD number DKFZp564) and hfkd2 (human fetal kidney; DKFZp566) libraries were generated using the Smart kit (Clontech), except that PCR was carried out with primers that contained uracil residues to permit directional cloning without restriction digestion and ligation, and were complementary with the pAMP1 (LifeTechnologies) cloning sites for directional cloning. The htes3 (human testes; DKFZp434), hute1 (human uterus; DKFZp586) and hmcfl (human mammary carcinoma; DKFZp727) libraries are conventional (Gubler, U., Hoffman, B.J., (1983), A simple and very efficient method for generating cDNA libraries. Gene 25, 263-269), size-selected cDNA libraries. They are cloned into pSPORT1 (LifeTechnologies) via a NotI site which is introduced during reverse transcription downstream of the oligo dT primer and a SalI site that is introduced by the ligation of a adapters. The human mammary carcinoma library was constructed fgrom MCF7 cells.

The cDNA sequences of this application were first identified among the sequences comprising various libraries. Technology has advanced considerably since the first cDNA libraries were made. Many small variations in both chemicals and machinery have been instituted over time, and these have improved both the efficiency and safety of the process. Although the cDNAs could be obtained using an older procedure, the procedure presented in this application is exemplary of one currently being used by persons skilled in the art. For the

purpose of providing an exemplary method, the mRNA isolation and cDNA library construction described here is for the MCF-7 library (DKFZp727) from which the clones named DKFZphmcf1_xxyyxx were obtained.

The human cell line MCF-7 was grown in DMEM supplemented with 10% fetal calf serum until confluency. 3 X 10^8 cells were harvested with a cell scraper in PBS. Cells were lysed in buffer containing 0.5 % NP-40 to leave the nuclei intact. The debris was pelleted by centrifugation at 15 000 x g for 10 minutes at 4 degrees Celsius. Proteins in the supernatant were degraded in presence of SDS and Proteinase K (30 minutes at 56 degrees Celsius). Precipitation of proteins was done in a Phenol/Chloroform extraction, RNA was precipitated from the aqueous phase with Na-acetate and Ethanol. Polyadenylated messages were isolated using Qiagen Oligotex (QIAGEN, Hilden Germany).

First strand cDNA synthesis was accomplished using an oligo (dT) primer which also contained an NotI restriction site. Second strand synthesis was performed using a combination of DNA polymerase I, *E. coli* ligase and RNase H, followed by the addition of a SalI adaptor to the blunt ended cDNA. The SalI adapted, double-stranded cDNA was then digested with NotI restriction enzyme, and fractionated by size on an agarose gel. DNA of the appropriate size was cut from the gel and cast into a second gel in a 90° angle. After electrophoresis in the second dimension, cDNA of the appropriate size was cut from the gel. The agarose block was broken down with help of gelase. The cDNA was purified with help of two phenol extractions and an ethanol precipitation. The cDNA was ligated into SalI/NotI pre-digested pSport1 vector (LifeTechnologies) and transformed into DH10B bacteria.

The libraries were arrayed into 384-well microtiter plates and spotted on high density nylon membranes for hybridization analysis. Filters and clones are available through the Resource Center. Whole plates were distributed to the sequencing partners of the consortium for systematic sequencing.

EXAMPLE II: Sequencing of cDNA Clones

All clones in the 384-well microtiter plates were sequenced from the 5' end. Sequencing was done preferentially using dye terminator chemistry (ABD or Amersham) on

ABI automated DNA sequencers (ABI 377, Applied Biosystems), one partner used EMBL prototype instruments (Arakis) mainly with dye primer chemistry.

The resulting expressed sequence tag (EST) sequences ("r1 ESTs" = sequenced from 5′-end) were analysed for:

a) the lack of identical matches with known genes.

For this, the EST-sequence was blasted against the cDNA consortiums own database and after that against public databases and (with BLASTn and BLASTx against EMBL/EMBLNEW and assembled ESTs, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings). ESTs which were identical to known genes in more than 100 bp, with less than 2 mismatches, were excluded from further analysis.

b) the presence of an open reading frame

Open reading frames (ORFs) were detected with an tool developed by Munich Information Center for Protein Sequences (MIPS) called ORF-map. ORF-map visualises potential start and stop-codons. If an ORF without a stop codon was detected in a r1-EST, the sequence was processed further.

c) the presence of GC rich sequences

A script developed by MIPS computed the GC-content of the r1-sequence, which should be >40%. Writing similar scripts is within the ordinary skill of one in bioinformatics.

d) the lack of repeat structures

Repeats such as Alu, Line or CA-repeats were detected by blasting (BLASTn and BLASTx, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings) against a repeat-database compiled by MIPS. If a repeat was present within the r1-sequence, the sequence were not processed further.

Novel clones that met all criteria were identified to the sequencers, who then performed 3'-end sequencing of these clones. The resulting 3' ESTs ("s1 ESTs" = sequenced from 3′-end) were checked for

a) the lack of matches with known genes in public databases, and sequences already generated by us.

This was done by blasting against EMBL/EMBLNEW and assembled EST (BLASTn and BLASTx, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings).

b) the presence of polyadenylation signals.

Again only clones matching the selection criteria were chosen to be sequenced completely by the sequencers. Clones were selected after the following criteria:

A very good ORF had at least one BLASTx match to other proteins. A "good ORF" should extend to the 3' end and be longer than ~40 codons. If the ORF started in the r1 sequence, in front of the potential start codon, there should not exist too many competing start codons in frame with the ORF start codon and the start should match the Kozak consensus ATG. If the EST sequence was to short to decide according to the potential ORF, and there were only a few or no start codons in the sequence the GC content of the Sequence should be greater than 40%. The r1 sequences needed not contain an polyA-tail at the 3' end. In addition, the results of the blasting against the assembled human ESTs could help in questionable cases to decide whether to stop or to continue. A hit against these ESTs was an indication to go further.

Clones passing the above-described screening were sequenced in full. Sequencing was done preferentially using dye terminator chemistry (ABD or Amersham) on ABI automated DNA sequencers (ABI 377, Applied Biosystems), one partner used EMBL prototype instruments (Arakis) mainly with dye primer chemistry. Primer walking (Strauss et al., 1986, Specific-primer-directed DNA sequencing. Anal Biochem. 154, 353-360) was the preferred sequencing strategy because of the lower redundancy possible compared to random shotgun (Messing, J., Crea, R., Seeburg, H.P. (1981) A system for shotgun DNA sequencing. Nucleic Acids Res. 9, 32-39) methods. Walking primers were generally designed using software (e.g. Haas, S., Vingron, M., Poustka, A., Wiemann, S. (1998) Primer design in large-scale sequencing. Nucleic Acids Res. 26, 3006-3012, Schwager, C., Wiemann, S., Ansorge, W. (1995) GeneSkipper: integrated software environment for DNA sequence assembly and

alignment. HUGO Genome Digest 2, 8-9) that permitted complete automation of this usually time consuming process and helped in the parallel processing of large numbers of clones.

EXAMPLE III: Bioinformatics analysis of full length cDNAs

Each sequence obtained was compared on nucleotide level in a stepwise manner to sequences in EMBL/EMBLNEW, EMBL-EST, EMBL-STS using the BLASTn algorithm. Basic Local Alignment Search Tool (BLAST, Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S. F. et al (1990) J Mol Biol 215:403-10) is used to search for local sequence alignments. BLAST produces alignments of both nucleotide (BLASTn) and amino acid sequences (BLASTp or BLASTx) to determine sequence similarity. BLAST is especially useful in determining exact matches or in identifying homologs, because of the local nature of the alignments. While it is useful for matches which do not contain gaps, it is inappropriate for performing motif-style searching. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment BLAST approach is to look threshold or cut off score set by the user. BLAST looks for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output. Parameter settings for the BLAST-operations (BLASTN 2.0a19MP-WashU) described were: EMBL-EMBLNEW: H=0 V=5 B=5 -filter seg; EMBL-EST: H=0 E=1e-10 B=500 V=500 -filter seg; EMBL-STS: H=0 V=5 B=5.

Search against EMBL/EMBLNEW was done to determine whether the cDNAs are already known, and also to find out whether the cDNAs are encoded by genomic sequences already sequenced and published/submitted to these databases.

Search against EMBL-EST was performed to get a first impression how abundant a particular cDNA would be and to get information on tissue specificity (so-called "electronic Northern-Blot", e.g. some of the cDNAs derived of the testis library show only hits to ESTs also derived of testis libraries).

The cDNA-sequences were blasted against EMBL-STS to determine STS-sequence-match to the cDNA, thus providing a mapping information to the new cDNA.

The potential protein-sequences were generated automatically by a script searching for the longest open reading frame (ORF) in each of the three forward frames with a minimum length of 90 codons. Next, the automatically generated ORFs were translated into protein sequences. These protein sequences were searched against the non redundant protein data set of PIR/SwissProt/Trembel/Tremblnew (BLASTP 2.0a19MP-WashU, parameter setting: V=7 B=7 H=0 -filter seg). If the script generated more than one ORF, one ORF was chosen manually by the annotater according to the degree of similarity to known proteins, the location of the ORF in the cDNA, the length, the amino acid composition and the content of Prosite-Motifs.

Additionally there was a BLASTx (BLASTX 2.0a19MP-WashU against non redundant protein database comprising PIR/SWISSPROT/TREMBL/TREMBLNEW; parameter-settings were: matrix/home/data/blast/matrix/aa/BLOSUM62 H=0 V=5 B=5 -filter seg) search to find potential frame shift in the complementary cds of the cDNAs and to identify unspliced or partly spliced cDNAs. The protein sequence was then transferred to the PEDANT system, in order to generate additional information on the new proteins. PEDANT (Protein Extraction, Description, and ANalysis Tool, Frishman, D. & Mewes, H.-W. (1997) PEDANTic genome analysis. Trends in Genetics , 13, 415-416) is a platform developed at the Munich Information Center for Protein Sequences (MIPS, Munich, Germany), which incorporates practically all bioinformatics methods important for the functional and structural characterisation of protein sequences. Computational methods used by PEDANT are:

FASTA

Very sensitive protein sequence database searches with estimates of statistical significance. Pearson W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. 183, 63-98.

BLAST2

Very sensitive protein sequence database searches with estimates of statistical significance. Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J. Basic local alignment search tool. Journal of Molecular Biology 215, 403-10.

PREDATOR

High-accuracy secondary structure prediction from single and multiple sequences. Frishman, D. and Argos, P. (1997) 75% accuracy in protein secondary structure prediction. Proteins, 27, 329-335. Frishman, D. and Argos, P.(1996) Incorporation of long-distance interactions in a secondary structure prediction algorithm. Prot. Eng. 9, 133-142.

STRIDE

Secondary structure assignment from atomic coordinates. Frishman, D. and Argos, P.(1995) Knowledge-based secondary structure assignment. Proteins 23, 566-579.

CLUSTALW

Multiple sequence alignment. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22:4673-4680.

TMAP

Transmembrane region prediction from multiply aligned sequences. Persson, B. and Argos, P. (1994) Prediction of transmembrane segments in proteins utilising multiple sequence alignments. J. Mol. Biol. 237, 182-192.

ALOM2

Transmembrane region prediction from single sequences. Klein, P., Kanehisa, M., and DeLisi, C. Prediction of protein function from sequence properties: A discriminant analysis of a database. Biochim. Biophys. Acta 787, 221-226 (1984). Version 2 by Dr. K. Nakai.

SIGNALP

Signal peptide prediction Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10, 1-6.

SEG

Detection of low complexity regions in protein sequences. Wootton, J.C., Federhen, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases. Computers & Chemistry 17, 149-163.

COILS

Detection of coiled coils. Lupas, A., M. Van Dyke, and J. Stock, "Predicting Coiled Coils from Protein Sequences." Science (1991) 252, 1162-1164.

PROSEARCH

Detection of PROSITE protein sequence patterns. Kolakowski L.F. Jr., Leunissen J.A.M., Smith J.E. (1992) ProSearch: fast searching of protein sequences with regular expression patterns related to protein structure and function. Biotechniques 13, 919-921.

BLIMPS

Similarity searches against a database of ungapped blocks. J.C. Wallace and Henikoff S., (1992) PATMAT: a searching and extraction program for sequence, pattern and block queries and databases, CABIOS 8, 249-254. Written by Bill Alford.

HMMER

Hidden Markov model software . Sonnhammer E.L.L., Eddy S.R., Durbin R. (1997) Pfam: A Comprehensive Database of Protein Families Based on Seed Alignments. Proteins 28, 405-420.

pI

Perl script that returns the amino acid composition, molecular weight, theoretical pI, and expected extinction coefficient of an amino acid sequence. By Fred Lindberg. The parameter-settings were as follows: known3d: score > 100; BLAST: E-value < 10; SCOP: <= 50 Alignments, E-Value < 0.0001; signalp: Y=0.7; untersucht vom N-Terminus her: 50 aa; funcat: E-value < 0.001; BLOCKS: <= 10 hits; BLIMPS: threshold 1100.0; COILS: threshold 0.95; SEG: threshold 20.0; BLAST in report: E-value < 0.001; PIR-KW, superfamilies, EC-Nummern in report: E-value < 0.00001; known3d in report: score > 120

The results of PEDANT analysis, together with the results of the similarity searches, constitute the basis for the structural and functional annotation of the cDNAs and the encoded proteins, as specified below.

EXAMPLE III: CELLULAR LOCALIZATIONS OF GFP-FUSION PROTEINS

Plasmids of cDNA-GFP fusions were transfected into mammalian tissue culture cells and allowed to express the proteins for up to 48 hours. Live cells were imaged at 24 hours and 48 hours after transfection and the localisations recorded. The chart, below, depicts the apparent final cellular localisations of 107 cDNA-GFP fusions.

In order to minimize the possibility of the GFP interfering with protein function and/or localization, two separate populations of cDNAs were generated encoding N-terminal or C-terminal GFP fusions. Clearly this appears to be a crucial strategy, since overall only 56% of the proteins localised to a specific compartment irrespective of the position of the GFP. In the instances where only one fusion localized, the complementary fusion either gave no expression or a nuclear and cytosolic staining - characteristic for GFP alone expression.

Each cDNA in turn was subjected to bioinformatic analysis. Where possible, the potential subcellular localisations of the expressed proteins were determined. This

information was then compared to the actual localisations determined from expression of the GFP-fusion proteins in mammalian cells.

DKFZphfbr2_16c16
----

group: Cell structure and motility

DKFZphfbr2_16c16.3 encodes a novel 586 amino acid protein with similarity to the human actin binding protein MAYVEN and Drosophila Kelch.

MAVEN is a novel actin binding protein predominantly expressed in brain. Drosophila kelch is involved in the maintenance of ring canal organization during oogenesis. The amino half of the protein including the BTB domain mediates dimerization, while the amino half might allow cross-linking of ring canal actin filaments, thus organising the inner rim cytoskeleton. The kelch repeat domain is necessary for ring canal localisation and believed to mediate an additional interaction, possibly with actin. The new protein shares the features of both proteins and therefore should be involved in the organisation of cyto skeleton binding to membrane proteins.

The new protein can find application in modulating/blocking of cyto skeleton-membrane protein interaction.

similarity to Drosophila kelch

complete cDNA, complete cds, EST hits
on genomic level partly encoded by AC005082 and AC006039

Sequenced by Qiagen

Locus: unknown

Insert length: 3028 bp
Poly A stretch at pos. 3004, polyadenylation signal at pos. 2984

```
   1 GGGGGCCCGG GGACGCAGCC CAGTTGGTAG CGTCGCTCCC TGAGCGTTTC
  51 TAAGGGGGCC GCCCGGCCCT GTCTTTCGGC AGTGGCCGAG CCACCGCCGC
 101 CTGCCGCGCG TTCCAGAGCT GGGCGCTGCA GCTGCACTGC CGATCGCCGT
 151 GTTTGGTCGA TAGAATCCCC AGTGTGCCCA GAGAGTGCGA CCCCTCGCCC
 201 GGCCCGGCGA GCCCCGGGCG TGAACCGAGC TGAGGGAGGA TGGCAGCCTC
 251 TGGGGTGGAG AAGAGCAGCA AGAAGAAGAC CGAGAAGAAA CTTGCTGCTC
 301 GGGAAGAAGC TAAATTGTTG GCGGGTTTCA TGGGCGTCAT GAATAACATG
 351 CGGAAACAGA AAACGTTGTG TGACGTGATC CTCATGGTCC AGGAAAGAAA
 401 GATACCTGCT CATCGTGTTG TTCTTGCTGC AGCCAGTCAT TTTTTTAACT
 451 TAATGTTCAC AACTAACATG CTTGAATCAA AGTCCTTTGA AGTAGAACTC
 501 AAAGATGCTG AACCTGATAT TATTGAACAA CTGGTGGAAT TTGCTTATAC
 551 TGCTAGAATT TCCGTGAATA GCAACAATGT TCAGTCTTTG TTGGATGCAG
 601 CAAACCAATA TCAGATTGAA CCTGTGAAGA AAATGTGTGT TGATTTTTTG
 651 AAAGAACAAG TTGATGCTTC AAATTGTCTT GGTATAAGTG TGCTAGCGGA
 701 GTGTCTAGAT TGTCCTGAAT TGAAAGCAAC TGCAGATGAC TTTATTCATC
 751 AGCACTTTAC TGAAGTTTAC AAAACTGATG AATTTCTTCA ACTTGATGTC
 801 AAGCGAGTAA CACATCTTCT CAACCAGGAC ACTCTGACTG TGAGAGCAGA
 851 GGATCAGGTT TATGATGCTG CAGTCAGGTG GTTGAAATAC GATGAGCCTA
 901 ATCGCCAGCC ATTTATGGTT GATATCCTTG CTAAAGTCAG GTTTCCTCTT
 951 ATATCAAAGA ATTTCTTAAG TAAAACGGTA CAAGCTGAAC CACTTATTCA
1001 AGACAATCCT GAATGCCTTA AGATGGTGAT AAGTGGAATG AGGTACCATC
1051 TACTGTCTCC AGAGGACCGA GAAGAACTTG TAGATGGCAC AAGACCTAGA
1101 AGAAAGAAAC ATGACTACCG CATAGCCCTA TTTGGAGGCT CTCAACCACA
1151 GTCTTGTAGA TATTTTAACC CAAAGGATTA TAGCTGGACA GACATCCGCT
1201 GCCCCTTTGA AAAACGAAGA GATGCAGCAT GCGTGTTTTG GGACAATGTA
1251 GTATACATTT TGGGAGGCTC TCAGCTTTTC CCAATAAAGC GAATGGACTG
1301 CTATAATGTA GTGAAGGATA GCTGGTATTC GAAACTGGGT CCTCCGACAC
1351 CTCGAGACAG CCTTGCTGCA TGTGCTGCAG AAGGCAAAAT TTATACATCT
1401 GGAGGTTCAG AAGTAGGAAA CTCAGCTCTG TATTTATTTG AGTGCTATGA
1451 TACGAGAACT GAAAGCTGGC ACACAAAGCC CAGCATGCTG ACCCAGCGCT
1501 GCAGCCATGG GATGGTGGAA GCCAATGGCC TAATCTATGT TTGTGGTGGA
1551 AGTTTAGGAA ACAATGTTTC AGGGAGAGTG CTTAATTCCT GTGAAGTTTA
1601 TGATCCTGCC ACAGAAACAT GGACTGAGCT GTGTCCAATG ATTGAAGCCA
1651 GGAAGAATCA TGGGCTGGTA TTTGTAAAAG ACAAGATATT TGCTGTGGGT
1701 GGTCAGAATG GTTTAGGTGG TCTGGACAAT GTGGAATATT ACGATATTAA
1751 GTTGAACGAA TGGAAGATGG TCTCACCAAT GCCATGGAAG GGTGTAACAG
1801 TGAAATGTGC AGCAGTTGGC TCTATAGTTT ATGTCTTGGC TGGTTTTCAG
```

```
2301 AGAAGATTGG CTCATCAGTG AAGCGCAGTA TCTTAGCTCT AGATTCTATT
2351 TTCATGCATC ACAGAAGTGC TATACGGTTA GGTCTGTTTG TGCTCAGTCA
2401 AGAACTAAGA AATAGTATGA ATTGTAAGTC AAGATGGGCA ACTCAGATGG
2451 AGCAGCTTAG TCTCACAGTT TGCTTGTCTA TTTATTTTAT TTAGTGCCAA
2501 ATGTATTCCA TTTTAAAAGT AAGCCAGAGT GAGTCAAGGC ATATACACAC
2551 TTTCTCACAA AACTTCCTAA ACAGATTTGG GGGTTTAATA TGTCCAACTC
2601 CTCATGAAAT ATATTCAATC CACTTAAATA TATTCCATCT TTTTAACATA
2651 AAATGTAAAG CTTAGCACCC ATCATTAATT TATGTCTCTG TTTTATCCAG
2701 TGGTTAAAAA AGGATTCTGC CTCTTTAGTC CTCACTGTTA AATAAAACCC
2751 AATCATAGTA AGTGATTAAC TAGCAAAAAG TAAAGCTATT TATAGCAAAT
2801 TTCTAGATCA TTAGAAAAGC ACTGGTAGTT GTACAATATC AGTGTTGACT
2851 TTGAACTTCT TTAACGAGAT CATGAATTCT TTTCCCTTAG CCAAAACATG
2901 AAATATTTAA CCTAGTTGTC TCTAAAAGTT TTGTAATCAT GAGTTAGATA
2951 TATGTCATCT CCTATTCATT GCTTTTATGT GATCAATAAA TCTTTTACAA
3001 ACCCAAAAGA AAAAAAAAAA AAAAAAA
```

BLAST Results

Entry AC005082 from database EMBL:
Homo sapiens clone RG271G13; HTGS phase 1, 7 unordered pieces.
Score = 6460, P = 0.0e+00, identities = 1292/1292
4 exons matching Bp 1180-3007

Entry AC006039 from database EMBL:
*** SEQUENCING IN PROGRESS *** Homo sapiens clone NH0319F03; HTGS phase 1, 3 unordered pieces.
Score = 1780, P = 2.0e-117, identities = 368/377
5 exons matching Bp 6-860

Entry HSG20603 from database EMBL:
human STS A005Y34.
Score = 670, P = 1.0e-23, identities = 134/134

Medline entries

93201592:
kelch encodes a component of intercellular bridges in Drosophila egg chambers.

97412177:
Drosophila kelch is an oligomeric ring canal actin organizer.

Peptide information for frame 3

ORF from 240 bp to 1997 bp; peptide length: 586
Category: strong similarity to known protein

```
  1 MAASGVEKSS KKKTEKKLAA REEAKLLAGF MGVMNNMRKQ KTLCDVILMV
 51 QERKIPAHRV VLAAASHFFN LMFTTNMLES KSFEVELKDA EPDIIEQLVE
101 FAYTARISVN SNNVQSLLDA ANQYQIEPVK KMCVDFLKEQ VDASNCLGIS
151 VLAECLDCPE LKATADDFIH QHFTEVYKTD EFLQLDVKRV THLLNQDTLT
201 VRAEDQVYDA AVRWLKYDEP NRQPFMVDIL AKVRFPLISK NFLSKTVQAE
251 PLIQDNPECL KMVISGMRYH LLSPEDREEL VDGTRPRRKK HDYRIALFGG
301 SQPQSCRYFN PKDYSWTDIR CPFEKRRDAA CVFWDNVVYI LGGSQLFPIK
351 RMDCYNVVKD SWYSKLGPPT PRDSLAACAA EGKIYTSGGS EVGNSALYLF
401 ECYDTRTESW HTKPSMLTQR CSHGMVEANG LIYVCGGSLG NNVSGRVLNS
451 CEVYDPATET WTELCPMIEA RKNHGLVFVK DKIFAVGGQN GLGGLDNVEY
501 YDIKLNEWKM VSPMPWKGVT VKCAAVGSIV YVLAGFQGVG RLGHILEYNT
551 ETDKWVANSK VRAFPVTSCL ICVVDTCGAN EETLET
```

BLASTP hits

Entry KELC_DROME from database SWISSPROT:
RING CANAL PROTEIN (KELCH PROTEIN).
Length = 689
Score = 816 (287.2 bits), Expect = 1.9e-81, P = 1.9e-81
Identities = 187/542 (34%), Positives = 290/542 (53%)

Entry AC004021_1 from database TREMBL:
WUGSC:H_DJ0186K10.1"; Human PAC clone DJ0186K10 from 5q31, complete sequence. Homo sapiens (human)
Length = 497

```
Score = 704 (247.8 bits), Expect = 1.4e-69, P = 1.4e-69
Identities = 163/483 (33%), Positives = 253/483 (52%)

Entry HSDKG12_1 from database TREMBL:
"KIAA0132";  Human mRNA for KIAA0132 gene, complete cds. Homo
sapiens (human)
Length = 624
Score = 692 (243.6 bits), Expect = 2.6e-68, P = 2.6e-68
Identities = 175/527 (33%), Positives = 272/527 (51%)

Entry A45773 from database PIR:
kelch protein, long form - fruit fly (Drosophila melanogaster)
Length = 1476
Score = 817 (287.6 bits), Expect = 1.7e-80, P = 1.7e-80
Identities = 189/549 (34%), Positives = 292/549 (53%)


          Alert BLASTP hits for DKFZphfbr2_16c16, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_16c16, frame 3
-----------------------------------------------

                    Report for DKFZphfbr2_16c16.3


[LENGTH]        586
[MW]            65992.06
[pI]            6.08
[HOMOL]         PIR:A45773 kelch protein, long form - fruit fly (Drosophila melanogaster) 5e-85

[BLOCKS]        BL00075D Dihydrofolate reductase proteins
[SCOP]          d1gog_3 2.46.1.1.1 (151-537) Galactose oxidase, central domai 6e-36
[PIRKW]         zinc finger 2e-11
[PIRKW]         DNA binding 9e-10
[PIRKW]         transcription factor 1e-06
[SUPFAM]        A55R protein middle region homology 1e-35
[SUPFAM]        POZ domain homology 1e-35
[SUPFAM]        vaccinia virus 59K HindIII-C protein 5e-15
[SUPFAM]        A55R protein 1e-35
[SUPFAM]        myxoma virus M9-R protein 2e-11
[SUPFAM]        A55R protein carboxyl-terminal homology 1e-35
[PROSITE]       CAMP_PHOSPHO_SITE      2
[PROSITE]       MYRISTYL       8
[PROSITE]       CK2_PHOSPHO_SITE       10
[PROSITE]       TYR_PHOSPHO_SITE       1
[PROSITE]       PKC_PHOSPHO_SITE       11
[PROSITE]       ASN_GLYCOSYLATION      1
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY     3.75 %

SEQ     MAASGVEKSSKKKTEKKLAAREEAKLLAGFMGVMNNMRKQKTLCDVILMVQERKIPAHRV
SEG     ......xxxxxxxxxxxxxxxxxxxxx.................................
PRD     ccceeeecccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeeccccchhhhe

SEQ     VLAAASHFFNLMFTTNMLESKSFEVELKDAEPDIIEQLVEFAYTARISVNSNNVQSLLDA
SEG     ............................................................
PRD     eccccccccccccccccchhhhheeeecccccchhhhhhhhhhhhheeeeccchhhhhhh

SEQ     ANQYQIEPVKKMCVDFLKEQVDASNCLGISVLAECLDCPELKATADDFIHQHFTEVYKTD
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ     EFLQLDVKRVTHLLNQDTLTVRAEDQVYDAAVRWLKYDEPNRQPFMVDILAKVRFPLISK
SEG     ............................................................
PRD     hhhchhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhccch

SEQ     NFLSKTVQAEPLIQDNPECLKMVISGMRYHLLSPEDREELVDGTRPRRKKHDYRIALFGG
SEG     ............................................................
PRD     hhhhhhhhhhcccccccchhhhhhhhhhhcccccccccccccccccccccceeeeeeeecc

SEQ     SQPQSCRYFNPKDYSWTDIRCPFEKRRDAACVFWDNVVYILGGSQLFPIKRMDCYNVVKD
SEG     ............................................................
PRD     ccccceeeccccccccccccccccccceeeeeeeeceeeeeeeccccccccccceeeecccccc

SEQ     SWYSKLGPPTPRDSLAACAAEGKIYTSGGSEVGNSALYLFECYDTRTESWHTKPSMLTQR
SEG     ............................................................
PRD     ccccccccccccccceeeeeccceeeeeeccccccccccceeeeeecccccccccccccccc
```

```
SEQ  CSHGMVEANGLIYVCGGSLGNNVSGRVLNSCEVYDPATETWTELCPMIEARKNHGLVFVK
SEG  ............................................................
PRD  ccceeeecceeeeeeccccccccccccccceeeeccccccccccccccccccccceeeeec

SEQ  DKIFAVGGQNGLGGLDNVEYYDIKLNEWKMVSPMPWKGVTVKCAAVGSIVYVLAGFQGVG
SEG  ............................................................
PRD  ceeeeccccccccccccceeeccccccccceeeccccccccccceeeeecccceeeeeccccccc

SEQ  RLGHILEYNTETDKWVANSKVRAFPVTSCLICVVDTCGANEETLET
SEG  ..............................................
PRD  ccccceeeccccccccccccccccccccccceeeeeeeeecccccccccccc
```

Prosite for DKFZphfbr2_16c16.3

PS00001	442->446	ASN_GLYCOSYLATION	PDOC00001
PS00004	11->15	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	188->192	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS00005	10->13	PKC_PHOSPHO_SITE	PDOC00005
PS00005	14->17	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	200->203	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	418->421	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00005	520->523	PKC_PHOSPHO_SITE	PDOC00005
PS00005	552->555	PKC_PHOSPHO_SITE	PDOC00005
PS00006	4->8	CK2_PHOSPHO_SITE	PDOC00006
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	315->319	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	405->409	CK2_PHOSPHO_SITE	PDOC00006
PS00006	460->464	CK2_PHOSPHO_SITE	PDOC00006
PS00006	550->554	CK2_PHOSPHO_SITE	PDOC00006
PS00007	202->209	TYR_PHOSPHO_SITE	PDOC00007
PS00008	5->11	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	389->395	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	436->442	MYRISTYL	PDOC00008
PS00008	440->446	MYRISTYL	PDOC00008
PS00008	487->493	MYRISTYL	PDOC00008
PS00008	493->499	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16c16.3)

DKFZphfbr2_16f21

group: brain derived

DKFZphfbr2_16f21 encodes a novel 208 amino acid protein with strong similarity to human zinc finger protein 216.

The novel protein shows strong similarity to the human zinc finger protein 216, but has no Zn finger.

PROSITE: Contains no Zinc finger; No informative BLAST results; no predictive prosite, pfam or SCOP motife

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to zinc finger protein 216

complete cDNA, complete cds, EST hits
start matches Kozak consensus ANNatgG,

Sequenced by Qiagen

Locus: unknown

Insert length: 1512 bp
Poly A stretch at pos. 1490, polyadenylation signal at pos. 1474

```
   1 GGGAGCAAGC AGGGGTTCGG CGGCATTACC TGTACCCATT CACCGGCGGC
  51 TACCGGCGGC GGCGCGTAGC GTGTCAGGCG GAGAGACCCG CCGCCAGGTG
 101 TGCAACTGAG GAACATGGCT CAAGAAACTA ATCACAGCCA AGTGCCTATG
 151 CTTTGTTCCA CTGGCTGTGG ATTTTATGGA AACCCTCGTA CAAATGGCAT
 201 GTGTTCAGTA TGCTATAAAG AACATCTTCA AAGACAGAAT AGTAGTAATG
 251 GTAGAATAAG CCCACCTGCA ACCTCTGTCA GTAGTCTGTC TGAATCTTTA
 301 CCAGTTCAAT GCACAGATGG CAGTGTGCCA GAAGCCCAGT CAGCATTAGA
 351 CTCTACATCT TCATCTATGC AGCCCAGCCC TGTATCAAAT CAGTCACTTT
 401 TATCAGAATC TGTAGCATCT TCTCAATTGG ACAGTACATC TGTGGACAAA
 451 GCAGTACCTG AAACAGAAGA TGTGCAGGCT TCAGTATCAG ACACAGCACA
 501 GCAGCCATCT GAAGAGCAAA GCAAGCCTCT TGAAAAACCG AAACAAAAAA
 551 AGAATCGCTG TTTCATGTGC AGGAAGAAAG TGGGACTTAC TGGGTTTGAA
 601 TGCCGGTGTG GAAATGTTTA CTGTGGTGTA CACCGTTACT CAGATGTACT
 651 CAATTGCTCT TACAATTACA AAGCCGATGC TGCTGAGAAA ATCAGAAAAG
 701 AAAATCCAGT AGTTGTTGGT GAAAAGATCC AAAAGATTTG AACTCCTGCT
 751 GGAATACAAA ATTCTTGAGC ATCTGCAAAC TAAAAATTGA CTTGAGGTTT
 801 TTTTTTTCCT AGTCATTGGG AATGTAGAGC AGTGTATCTT GCATGTCATC
 851 GGAAGAATAG ATTTTTGTTT TGGTTTTGTT TTGAAAATGA CTCTGAACAT
 901 TTATTTCCAT TGCAATTTCT GTGGCTGAGG AGACTTAAAC TTTACAAGTA
 951 TTATCCTTTT AAGATCATTT TAATTTTAGT TGAGTGCAGA GGGCTTTTAT
1001 AACAAACGTG CAGAAATTTT GGAGGGCTGT GATTTTTCCA GTATTAAACA
1051 TGCATGCATT AATCTTGCAG TTTATTTTCT CATTATGTAT GTATATATCG
1101 CTTTTCTCTG CAGCACGATT TCTCTTTTGA TAATGCCCTT TAGGGCACAA
1151 CTAGTTATCA GTAACTGAAT GTATCTTAAT CATTATGGCT GCTTCTGTTT
1201 TTTCATTAAC AAAGGTTATT CATATGTTAG CATATAGTTT CTTTGCACCC
1251 ACTATTTATG TCTGAATCAT TTGTCACAAG AGAGTGTGTG CTGATGAGAT
1301 TGTAAGTTTG TGTGTTTAAA CTTTTTTTTG AGCGAGGGAA GAAAAAGCTG
1351 TATGCATTTC ATTGCTGTCT ACAGGTTTCT TTCAGATTAT GTTCATGGGT
1401 TTGTGTGTAT ACAATATGAA GAATGATCTG AAGTAATTGT GCTGTATTTA
1451 TGTTTATTCA CCAGTCTTTG ATTAAATAAA AAGGAAAACC AGAAAAAAAA
1501 AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 115 bp to 738 bp; peptide length: 208
Category: strong similarity to known protein

```
  1 MAQETNHSQV PMLCSTGCGF YGNPRTNGMC SVCYKEHLQR QNSSNGRISP
 51 PATSVSSLSE SLPVQCTDGS VPEAQSALDS TSSSMQPSPV SNQSLLSESV
101 ASSQLDSTSV DKAVPETEDV QASVSDTAQQ PSEEQSKPLE KPKQKKNRCF
151 MCRKKVGLTG FECRCGNVYC GVHRYSDVLN CSYNYKADAA EKIRKENPVV
201 VGEKIQKI
```

BLASTP hits

Entry ATF7H19_1 from database TREMBLNEW:
gene: "F7H19.10"; product: "putative protein"; Arabidopsis thaliana DNA chromosome 4, BAC clone F7H19 (ESSAII project) >TREMBL:ATT12H17_21
gene: "T12H17.210"; product: "predicted protein"; Arabidopsis thaliana DNA chromosome 4, BAC clone T12H17 (ESSAII project)
Score = 206, P = 2.1e-24, identities = 51/146, positives = 77/146

Entry PVPVPR3A_1 from database TREMBL:
gene: "PVPR3"; P.vulgaris PVPR3 protein mRNA, complete cds.
Score = 237, P = 4.9e-20, identities = 50/136, positives = 73/136

Entry AF062072_1 from database TREMBL:
gene: "ZNF216"; product: "zinc finger protein 216"; Homo sapiens zinc finger protein 216 (ZNF216) gene, complete cds.
Score = 591, P = 1.6e-57, identities = 124/215, positives = 147/215

Alert BLASTP hits for DKFZphfbr2_16f21, frame 1

TREMBL:AF062071_1 product: "zinc finger protein ZNF216"; Mus musculus zinc finger protein ZNF216 mRNA, complete cds., N = 1, Score = 590, P = 2.1e-57

TREMBLNEW:AB001773_1 gene: "pem-6"; product: "PEM-6"; Ciona savignyi pem-6 (posterior end mark 6) mRNA, complete cds., N = 1, Score = 421, P = 1.7e-39

```
>TREMBL:AF062071_1 product: "zinc finger protein ZNF216"; Mus musculus zinc
     finger protein ZNF216 mRNA, complete cds.
            Length = 213

  HSPs:

 Score = 590 (88.5 bits), Expect = 2.1e-57, P = 2.1e-57
 Identities = 123/213 (57%), Positives = 146/213 (68%)

Query:     1 MAQETNHSQVPMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQNSSNGRISPPAT---SVSS 57
             MAQETN +  PMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQ +S GR+SP  T    S S
Sbjct:     1 MAQETNQTPGPMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQQNS-GRMSPMGTASGSNSP 59

Query:    58 LSESLPVQCTDGSVPEAQSALDSTSSSMQPSPVSNQSLLSE--SVASSQLDSTSVDKAVP 115
             S+S  VQ  D +    + A  STS    +  PV+   +  +    ++ S+ D  +  K
Sbjct:    60 TSDSASVQRADAGLNNCEGAAGSTSEKSRNVPVAALPVTQQMTEMSISREDKITTPKT-E 118

Query:   116 ETEDVQASVSDTAQQPSEEQS--KPLEKPKQKKNRCFMCRKKVGLTGFECRCGNVYCGVH 173
             +E V    S +  QPS  QS  K  E PK KKNRCFMCRKKVGLTGF+CRCGN++CG+H
Sbjct:   119 VSEPVVTQPSPSVSQPSSSQSEEKAPELPKPKKNRCFMCRKKVGLTGFDCRCGNLFCGLH 178

Query:   174 RYSDVLNCSYNYKADAAEKIRKENPVVVGEKIQKI 208
             RYSD  NC Y+YKA+AA KIRKENPVVV EKIQ+I
Sbjct:   179 RYSDKHNCPYDYKAEAAAKIRKENPVVVAEKIQRI 213
```

Pedant information for DKFZphfbr2_16f21, frame 1

Report for DKFZphfbr2_16f21.1

[LENGTH] 208
[MW] 22541.23
[pI] 6.80
[HOMOL] TREMBL:AF062072_1 gene: "ZNF216"; product: "zinc finger protein 216"; Homo sapiens zinc finger protein 216 (ZNF216) gene, complete cds. 9e-57
[PIRKW] zinc 8e-13
[PIRKW] zinc finger 8e-13

[PIRKW]	fusion protein 8e-13	
[SUPFAM]	unassigned ubiquitin-related proteins 8e-13	
[SUPFAM]	ubiquitin homology 8e-13	
[PROSITE]	MYRISTYL	2
[PROSITE]	CK2_PHOSPHO_SITE	7
[PROSITE]	ASN_GLYCOSYLATION	4
[KW]	Irregular	
[KW]	LOW_COMPLEXITY	7.21 %

```
SEQ   MAQETNHSQVPMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQNSSNGRISPPATSVSSLSE
SEG   ............................................................
PRD   ccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccccccccc

SEQ   SLPVQCTDGSVPEAQSALDSTSSSMQPSPVSNQSLLSESVASSQLDSTSVDKAVPETEDV
SEG   ...................xxxxxxxxxxxxxxx..........................
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   QASVSDTAQQPSEEQSKPLEKPKQKKNRCFMCRKKVGLTGFECRCGNVYCGVHRYSDVLN
SEG   ............................................................
PRD   cccccccccccccccccccccccccccccceeccccccccceeecccccccccccccccc

SEQ   CSYNYKADAAEKIRKENPVVVGEKIQKI
SEG   ............................
PRD   ccchhhhhhhhhhhhhcccccccccccc
```

Prosite for DKFZphfbr2_16f21.1

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	42->46	ASN_GLYCOSYLATION	PDOC00001
PS00001	92->96	ASN_GLYCOSYLATION	PDOC00001
PS00001	180->184	ASN_GLYCOSYLATION	PDOC00001
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	70->74	CK2_PHOSPHO_SITE	PDOC00006
PS00006	76->80	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	159->163	CK2_PHOSPHO_SITE	PDOC00006
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	166->172	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16f21.1)

DKFZphfbr2_16g18
----

group: cell cycle

DKFZphfbr2_16g18.3 encodes a novel 984 amino acid protein with similarity to centromeric proteins of yeasts.

The novel protein shows similarity to S. pombe SPAC17A5.07c and the S. cerevisiae Smt4p suppressor of MIF2 gene. MIF2 encodes a centromeric protein with homology to the mammalian centromeric protein CENP-C. Mutations in MIF2 stabilise dicentric minichromosomes and confer high instability to chromosomes that bear a cis-acting mutation in element I of the yeast centromeric DNA (CDEI). Therefore the new protein should be involved in centromer organisation, too.

The new protein can find application in modulating/blocking the cell cycle and influencing the behavior of chromosomes, both natural and artificial in eukaryotic cells.

similarity to KIAA0797 and yeast Smt4p

complete cDNA, complete cds, EST hits
the yeast Smt4 protein seems to be involved in centromer function
and microtuble organisation

Sequenced by Qiagen

Locus: unknown

Insert length: 4826 bp
Poly A stretch at pos. 4756, polyadenylation signal at pos. 4736

```
   1 GGGTCGAGGT CGACGGTATC GATAAGTTTT TTTTTTTTTT TTTTTTTTTT
  51 TTTTCCTTTC CCCTCCCCCT CCCTCTCCAA GCCGGAGGGG TCCTGAGGTG
 101 ACAGCGCCTG CAACTGAAAT TTCAGCAGCG GGAGAAGATG GACAAGAGAA
 151 AGCTCGGGCG ACGGCCATCT TCATCCGAAA TCATCACAGA AGGAAAAAGG
 201 AAAAAGTCAT CTTCTGATTT ATCGGAGATA AGAAAGATGT TAAATGCAAA
 251 ACCAGAGGAT GTCCATGTTC AATCACCACT GTCCAAATTC AGAAGCTCAG
 301 AACGCTGGAC TCTCCCTTTG CAGTGGGAAA GAAGCCTAAG GAATAAAGTC
 351 ATCTCTCTAG ACCATAAAAA TAAAAAACAT ATCCGAGGGT GTCCTGTTAC
 401 TTCCAGGTCA TCACCAGAAA GGATACCCAG AGTTATATTG ACGAATGTCC
 451 TGGGAACGGA GTTAGGAAGA AAATACATAA GGACCCCACC TGTAACTGAG
 501 GGAAGTTTGA GTGATACAGA CAACTTGCAA TCAGAGCAAC TTTCTTCATC
 551 ATCTGATGGC AGCCTAGAAT CTTATCAAAA TCTAAACCCT CACAAGAGCT
 601 GTTATTTATC TGAAAGGGGC TCACAACGAA GTAAGACAGT AGATGACAAT
 651 TCTGCAAAGC AGACTGCGCA CAATAAAGAA AAACGAAGAA AGGATGATGG
 701 CATTTCTCTT TTAATATCTG ATACTCAGCC TGAAGACCTT AACAGTGGAA
 751 GTAGAGGTTG TGATCATCTC GAACAGGAAA GCAGAAACAA GGATGTTAAA
 801 TATTCTGATT CAAAAGTGGA ACTCACTCTG ATTTCCAGGA AGACAAAGAG
 851 AAGGCTTAGA AATAATTTAC CTGATTCTCA ATATTGTACT TCTTTGGATA
 901 AGTCAACAGA ACAGACAAAA AAACAAGAAG ATGACTCAAC AATATCCACT
 951 GAGTTTGAAA GGCCAAGTGA AAACTATCAT CAGGATCCAA AACTGCCTGA
1001 AGAAATTACA ACTAAACCTA CAAAAAGTGA TTTTACTAAG CTATCCTCAC
1051 TTAACAGTCA GGAGTTGACT TTGAGTAATG CCACCAAAAG TGCCTCTGCC
1101 GGTTCAACCA CTGAAACCGT TGAGTACTCT AATTCCATTG ATATTGTGGG
1151 GATTTCTTCC CTGGTTGAGA AGGATGAGAA TGAGTTGAAT ACCATAGAAA
1201 AGCCTATTCT AAGAGGACAT AATGAAGGGA ACCAATCACT GATCTCAGCT
1251 GAACCAATTG TTGTTTCCAG TGATGAAGAA GGACCTGTTG AACATAAAAG
1301 TTCAGAAATT CTTAAGTTAC AATCTAAGCA AGACCGTGAG ACAACTAATG
1351 AAAATGAGAG TACTTCTGAA TCAGCATTGT TAGAACTACC ATTGATTACA
1401 TGTGAATCTG TACAGATGTC ATCTGAATTA TGCCCATATA ATCCTGTCAT
1451 GGAGAACATT TCCAGTATTA TGCCTAGTAA TGAGATGGAT CTACAACTGG
1501 ATTTTATATT TACTTCTGTT TATATTGGTA AAATAAAAGG AGCTTCTAAA
1551 GGTTGTGTTA CAATCACAAA AAAATATATT AAGATCCCAT TTCAAGTGTC
1601 CCTGAATGAG ATTTCATTGC TAGTGGATAC CACACATTTA AAGCGGTTTG
1651 GGTTATGGAA AAGTAAGGAT GATAATCACA GTAAAAGGAG TCATGCTATT
1701 CTTTTCTTCT GGGTCTCTTC AGATTATCTT CAAGAGATTC AGACCCAATT
1751 AGAACACTCT GTATTAAGCC AGCAATCAAA ATCTAGTGAA TTCATTTTCC
1801 TTGAACTACA CAATCCTGTT TCACAGAGAG AAGAATTGAA GCTGAAAGAT
1851 ATTATGACGG AAATAAGTAT AATCAGTGGA GAATTAGAGC TTTCTTACCC
1901 GTTGTCTTGG GTTCAGGCAT TTCCTTTGTT TCAGAACCTC TCTTCAAAAG
1951 AAAGTTCTTT TATTCATTAT TACTGTGTTT CAACTTGTTC TTTCCCTGCT
2001 GGTGTTGCTG TTGCTGAAGA AATGAAGCTG AAATCAGTAT CTCAGCCCTC
2051 AAACACAGAT GCGGCCAAGC CTACTTACAC CTTCCTGCAG AAGCAAAGTA
2101 GCGGTTGCTA CTCCCTTTCT ATTACATCTA ATCCAGATGA AGAATGGCGG
2151 GAAGTCAGGC ACACTGGACT TGTTCAGAAG TTGATTGTAT ATCCTCCACC
2201 ACCTACTAAG GGGGGATTGG GAGTAACTAA TGAAGATCTG GAGTGTTTAG
2251 AAGAAGGAGA GTTTCTTAAT GATGTAATCA TTGATTTTTA CCTTAAGTAT
2301 CTTATATTGG AGAAGGCATC AGATGAACTT GTTGAACGAA GTCACATTTT
```

```
2351 TAGTAGCTTT TTCTATAAAT GCTTGACAAG AAAGGAAAAT AATTTAACAG
2401 AAGATAATCC AAATCTTTCA ATGGCACAGA GAAGACATAA AAGAGTAAGA
2451 ACATGGACTC GTCACATAAA CATTTTTAAT AAAGATTACA TCTTTGTACC
2501 TGTAAATGAG TCGTCTCACT GGTATCTCGC AGTCATTTGT TTTCCATGGT
2551 TAGAAGAAGC TGTGTATGAA GATTTTCCAC AAACTGTATC CCAGCAGTCC
2601 CAGGCTCAGC AGTCCCAAAG TGACAACAAA ACAATAGATA ATGATCTACG
2651 TACTACTTCG ACACTGTCTT TGAGTGCAGA GGATTCCCAA AGTACCGAGT
2701 CGAATATGTC AGTACCAAAG AAAATGTGTA AAAGGCCATG TATTCTTATA
2751 CTAGACTCCT TGAAAGCTGC TTCTGTACGA AACACAGTTC AGAATTTACG
2801 AGAGTATTTA GAGGTAGAGT GGGAAGTTAA ACTAAAAACT CATCGTCAAT
2851 TCAGCAAAAC AAACATGGTG GATCTATGCC CTAAAGTTCC TAAACAGGAC
2901 AATAGCAGTG ATTGTGGAGT ATATTTATTG CAGTATGTGG AAAGCTTCTT
2951 CAAGGATCCT ATTGTTAACT TTGAACTTCC AATTCATTTG GAGAAGTGGT
3001 TTCCTCGTCA TGTAATAAAG ACCAAACGGG AAGATATTCG AGAGCTCATC
3051 TTGAAACTTC ATTTACAGCA ACAGAAGGGC AGCAGTAGCT AGTTAATCTG
3101 TACAAACATG ACACAGATGT TCTCTAAGAT TACTGGAAAG CCCCTTACCA
3151 GCATTTGTGT TAGCCAGCTC ACAGAGAAGA AAATAACTTG CAGTAGTTTT
3201 ATAATAAGTC ATTGGAACAT TATTTAAAAT ATGTAGGACA CATTATTAGA
3251 ATTGTTGGGA TCTCATAGAT GGAATGGGAA TGGGGGTGAT ATAGATAAAC
3301 TTACTAGATA TAAATTAAAA TTTTATAAAT ATTTCATATT TTTCTGAGTA
3351 AATATGATTG GATTATGCAA CAGCATATGT AATATGGGAA TGTTTTGTAG
3401 ATAATAAAAC TTACATGATC TGTACTTCCA CGTGACTGGG TGCTGAGGGG
3451 AGTTAAAGCC TCCCTGGTGC CAGCCCCAGT GCTTGTCAAA TTTGCTGACA
3501 GGTCACATCA TATTGTAATT CTATTCTTTG CAGCTCAAGC ATGCAGTATG
3551 AATACTGTGT ATTTTTTAAA AAAATAATTT AGTATCAAGG CTTCAGAAAA
3601 TGCCATTTAC GGCATCCCTT CTGTATGTAA CAAAAAGACA TTCATAATGT
3651 TAGGAAGATG ATAAAAATTC GCTCTTTTAA AGTGCAGCTT ATTATTCTCA
3701 ATTGCTAAAT ACGATTACTC TGCTTTTTTT TTTTCATTTC TTTTGATGTC
3751 ATATGTGAGT ATCTTATAAT TTAGTTCATT TGTTCAGGGT AAAATTTGAA
3801 ACAAAAAATT TTACCTGTGC AAAATAGTTT TTTAAAAATT ATACATGTAG
3851 CTCAACTTGA GGTACTGCTA TATAAATATT CACTCACATT ATCACGGAAT
3901 TTATGTATAG TTTCTCTAAT ATAGAAGATA AAATTGGTGT CCTCATAACT
3951 TTAACAAAGA AAACCCTCAG TCCTATTTAT TAATGGGTAG AATTAAATAT
4001 ATAATTTTAT AGCTCAGTTT ACCCAGTATT CATCTGCAAA GCCAGATTGC
4051 TCTCATTGCT TTTATATTTT TAAATTGTAG CTTTTAGAGA CCTATGATCC
4101 TCATGGAACT TAATTTTTTA TTAAATATTC AGGTAACAGT TCTGAATTCA
4151 TGTGATAATG GTGGCATTAT ATATGATTAA ACACTTCAGA ACTTTCTAAT
4201 GTTATCAGGA GTATTTTGAG GGAGATATGA TTATATTGTA TTTTCTCAGA
4251 TAAGAAAAAT GTTTTTTAAC AATATTATTT TAATCTGTTT TAAGCATCTC
4301 TTAGATTTAC ATTATAACTA CATAAAGCAG TGAAGCAAAG GCAAATTAAG
4351 ATAAAGCTAG AAAGTCTGAA CATTTTATTT CAAAATCATA CGAATCGGGG
4401 TCAGTTAAGC CTCAGTATTC TTAGCTTTTG TTGATTTTGG CACTATCTTT
4451 ATATTATTAA ATATATTTGT TGTTTGGATA TTTCATATAA AGATGGCTAT
4501 AATTACATAT TTCATTCCCA ATTTGTGTGT GTTGGGGGGT ACTTTTAAAG
4551 GTGACTATTG TTTTGTACAT CTAATTTTGG GAAACCAAGT CTATAAGACA
4601 TCTTGTGATT TCTTAATGTT TTTGTTTGTA TGTTTTTCAA AGATATCACT
4651 GTCCTTTATC ATGTTTTGAA GATTGTTTAA AATTCATTTT CCTAAATTAA
4701 TGTGCAAGTA ATGTTTTGAG GATATCGGTG TTTTATATTA AACATATTTC
4751 CAATTCAAAA AAAAAAAAAA AAAAACTTAT CGATACCGTC GACCTCGATG
4801 ATGATGATGA TGATGATGAT GTCGAC
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 138 bp to 3089 bp; peptide length: 984
Category: similarity to known protein

```
  1 MDKRKLGRRP SSSEIITEGK RKKSSSDLSE IRKMLNAKPE DVHVQSPLSK
 51 FRSSERWTLP LQWERSLRNK VISLDHKNKK HIRGCPVTSR SSPERIPRVI
101 LTNVLGTELG RKYIRTPPVT EGSLSDTDNL QSEQLSSSSD GSLESYQNLN
151 PHKSCYLSER GSQRSKTVDD NSAKQTAHNK EKRRKDDGIS LLISDTQPED
201 LNSGSRGCDH LEQESRNKDV KYSDSKVELT LISRKTKRRL RNNLPDSQYC
251 TSLDKSTEQT KKQEDDSTIS TEFERPSENY HQDPKLPEEI TTKPTKSDFT
301 KLSSLNSQEL TLSNATKSAS AGSTTETVEY SNSIDIVGIS SLVEKDENEL
351 NTIEKPILRG HNEGNQSLIS AEPIVVSSDE EGPVEHKSSE ILKLQSKQDR
401 ETTNENESTS ESALLELPLI TCESVQMSSE LCPYNPVMEN ISSIMPSNEM
451 DLQLDFIFTS VYIGKIKGAS KGCVTITKKY IKIPFQVSLN EISLLVDTTH
```

```
501 LKRFGLWKSK DDNHSKRSHA ILFFWVSSDY LQEIQTQLEH SVLSQQSKSS
551 EFIFLELHNP VSQREELKLK DIMTEISIIS GELELSYPLS WVQAFPLFQN
601 LSSKESSFIH YYCVSTCSFP AGVAVAEEMK LKSVSQPSNT DAAKPTYTFL
651 QKQSSGCYSL SITSNPDEEW REVRHTGLVQ KLIVYPPPPT KGGLGVTNED
701 LECLEEGEFL NDVIIDFYLK YLILEKASDE LVERSHIFSS FFYKCLTRKE
751 NNLTEDNPNL SMAQRRHKRV RTWTRHINIF NKDYIFVPVN ESSHWYLAVI
801 CFPWLEEAVY EDFPQTVSQQ SQAQQSQSDN KTIDNDLRTT STLSLSAEDS
851 QSTESNMSVP KKMCKRPCIL ILDSLKAASV RNTVQNLREY LEVEWEVKLK
901 THRQFSKTNM VDLCPKVPKQ DNSSDCGVYL LQYVESFFKD PIVNFELPIH
951 LEKWFPRHVI KTKREDIREL ILKLHLQQQK GSSS
```

BLASTP hits

Entry SPAC17A5_7 from database TREMBL:
"SPAC17A5.07c"; product: "hypothetical protein"; S.pombe chromosome I cosmid c17A5. Schizosaccharomyces pombe (fission yeast)
Length = 652
Score = 275 (96.8 bits), Expect = 1.9e-29, Sum P(3) = 1.9e-29
Identities = 56/120 (46%), Positives = 78/120 (65%)

Entry S49947 from database PIR:
SMT4 protein - yeast (Saccharomyces cerevisiae)
Length = 1034
Score = 163 (57.4 bits), Expect = 4.6e-16, Sum P(3) = 4.6e-16
Identities = 46/159 (28%), Positives = 76/159 (47%)

Entry YQG6_CAEEL from database SWISSPROT:
HYPOTHETICAL 35.7 KD PROTEIN C41C4.6 IN CHROMOSOME II.
Length = 342
Score = 162 (57.0 bits), Expect = 6.1e-13, Sum P(3) = 6.1e-13
Identities = 37/119 (31%), Positives = 62/119 (52%)

Entry AB018340_1 from database TREMBL:
gene: "KIAA0797"; product: "KIAA0797 protein"; Homo sapiens mRNA for KIAA0797 protein, partial cds.
Score = 540, P = 1.9e-50, identities = 120/243, positives = 155/243

Alert BLASTP hits for DKFZphfbr2_16g18, frame 3

TREMBL:ATT16L1_11 gene: "T16L1.110"; product: "putative protein"; Arabidopsis thaliana DNA chromosome 4, BAC clone T16L1 (ESSAII project), N = 2, Score = 239, P = 2.1e-18

```
>TREMBL:ATT16L1_11 gene: "T16L1.110"; product: "putative protein";
     Arabidopsis thaliana DNA chromosome 4, BAC clone T16L1 (ESSAII project)
          Length = 710

  HSPs:

 Score = 239 (35.9 bits), Expect = 2.1e-18, Sum P(2) = 2.1e-18
 Identities = 51/135 (37%), Positives = 78/135 (57%)

Query:   683 IVYPPPPTKGGLGVTNEDLECLEEGEFLNDVIIDFYLKYLILEKASDELVERSHIFSSFF 742
             +VYP       + V  +D+E L+    F+ND IIDFY+KYL    + S +    R H F+ FF
Sbjct:   176 LVYPQGEPDAVV-VRKQDIELLKPRRFINDTIIDFYIKYL-KNRISPKERGRFHFFNCFF 233

Query:   743 YKCLTRKENNLTEDNPNLSMAQRRHKRVRTWTRHINIFNKDYIFVPVNESSHWYLAVICF 802
             +     RK  NL  +  P+    +   ++RV+ WT+++++F KDYIF+P+N S HW L +IC
Sbjct:   234 F----RKLANLDKGTPSTCGGREAYQRVQKWTKNVDLFEKDYIFIPINCSFHWSLVIICH 289

Query:   803 PWLEEAVYEDFPQTV 817
             P      + + PQ V
Sbjct:   290 PGELVPSHVENPQRV 304

 Score = 70 (10.5 bits), Expect = 2.1e-18, Sum P(2) = 2.1e-18
 Identities = 13/28 (46%), Positives = 15/28 (53%)

Query:   948 PIHLEKWFPRHVIKTKREDIRELILKLH 975
             P HL  WFP      KR +I EL+  LH
Sbjct:   403 PSHLRNWFPAKEASLKRRNILELLYNLH 430
```

Pedant information for DKFZphfbr2_16g18, frame 3

Report for DKFZphfbr2_16g18.3

```
[LENGTH]        984
[MW]            112265.80
[pI]            6.13
[HOMOL]         TREMBL:AB018340_1 gene: "KIAA0797"; product: "KIAA0797 protein"; Homo sapiens
mRNA for KIAA0797 protein, partial cds. 8e-53
[FUNCAT]        03.22 cell cycle control and mitosis [S. cerevisiae, YIL031w] 9e-17
[FUNCAT]        99 unclassified proteins     [S. cerevisiae, YPL020c] 4e-06
[BLOCKS]        BL00494C Bacterial luciferase subunits proteins
[PROSITE]       AMIDATION       3
[PROSITE]       MYRISTYL        9
[PROSITE]       CAMP_PHOSPHO_SITE       2
[PROSITE]       CK2_PHOSPHO_SITE        30
[PROSITE]       TYR_PHOSPHO_SITE        1
[PROSITE]       PKC_PHOSPHO_SITE        19
[PROSITE]       ASN_GLYCOSYLATION       12
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY     4.47 %
```

```
SEQ     MDKRKLGRRPSSSEIITEGKRKKSSSDLSEIRKMLNAKPEDVHVQSPLSKFRSSERWTLP
SEG     ............................................................
PRD     ccccceeeccccceeeeecccccccccchhhhhhhhhcccccccccccccccccccchh

SEQ     LQWERSLRNKVISLDHKNKKHIRGCPVTSRSSPERIPRVILTNVLGTELGRKYIRTPPVT
SEG     ............................................................
PRD     hhhhhhhhheeeecccccceeecccccccccccccceeeeeeeeecccceeeccceeecccc

SEQ     EGSLSDTDNLQSEQLSSSSDGSLESYQNLNPHKSCYLSERGSQRSKTVDDNSAKQTAHNK
SEG     .........xxxxxxxxxxxxxxxx...................................
PRD     ccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhh

SEQ     EKRRKDDGISLLISDTQPEDLNSGSRGCDHLEQESRNKDVKYSDSKVELTLISRKTKRRL
SEG     ............................................................
PRD     hhhhcccceeeeeccccccccccccccccccccccccccccccccceeeeeehhhhhh

SEQ     RNNLPDSQYCTSLDKSTEQTKKQEDDSTISTEFERPSENYHQDPKLPEEITTKPTKSDFT
SEG     ............................................................
PRD     hccccccccccccccccchhhhcccccccccccccccccccccccccccccccccccccc

SEQ     KLSSLNSQELTLSNATKSASAGSTTETVEYSNSIDIVGISSLVEKDENELNTIEKPILRG
SEG     ............................................................
PRD     cccccccccceeehhhhhhhcccccceeeecccceeeceeecccchhhhhhhhhhcccccc

SEQ     HNEGNQSLISAEPIVVSSDEEGPVEHKSSEILKLQSKQDRETTNENESTSESALLELPLI
SEG     ........................................xxxxxxxxxxxxxxxxx...
PRD     cccccceeeecceeeeeccccccccccchhhhhhhhhhhhhcccccccchhhhhccccce

SEQ     TCESVQMSSELCPYNPVMENISSIMPSNEMDLQLDFIFTSVYIGKIKGASKGCVTITKKY
SEG     ............................................................
PRD     eeccccccccccccccccccceeeccccchhhhhheeeeeeeeeeeecccccceeeeeeee

SEQ     IKIPFQVSLNEISLLVDTTHLKRFGLWKSKDDNHSKRSHAILFFWVSSDYLQEIQTQLEH
SEG     ............................................................
PRD     eeeecccccceeeeeeeecccceeeeeeeeecccccccccceeeeeeeeecchhhhhhhhhh

SEQ     SVLSQQSKSSEFIFLELHNPVSQREELKLKDIMTEISIISGELELSYPLSWVQAFPLFQN
SEG     ............................................................
PRD     hhhhcccccceeeeeeeeeccccccchhhhhhhhhheeeeeccceeeecccceeeeeeceeec

SEQ     LSSKESSFIHYYCVSTCSFPAGVAVAEEMKLKSVSQPSNTDAAKPTYTFLQKQSSGCYSL
SEG     ............................................................
PRD     cccccccccceeeeecccccccchhhhhhhhhhhcccccccccccccccceeeecccccccce

SEQ     SITSNPDEEWREVRHTGLVQKLIVYPPPPTKGGLGVTNEDLECLEEGEFLNDVIIDFYLK
SEG     ............................................................
PRD     eeccccccceeeeeecccceeeeeeecccccccccccccchhhhhhhhcchhhhhhhhhh

SEQ     YLILEKASDELVERSHIFSSFFYKCLTRKENNLTEDNPNLSMAQRRHKRVRTWTRHINIF
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccchhhhhhhhhhhhhhhhhhhc

SEQ     NKDYIFVPVNESSHWYLAVICFPWLEEAVYEDFPQTVSQQSQAQQSQSDNKTIDNDLRTT
SEG     .....................................xxxxxxxxxxx............
PRD     cceeeeeccccccceeeeeeeccchhhhhhhccccchhhhhhhhhhcccccccccccccc

SEQ     STLSLSAEDSQSTESNMSVPKKMCKRPCILILDSLKAASVRNTVQNLREYLEVEWEVKLK
SEG     ............................................................
PRD     cceeeeecccccceeeccccccccccccceeeeeccccccccchhhhhhhhhhhhhhhhhh

SEQ     THRQFSKTNMVDLCPKVPKQDNSSDCGVYLLQYVESFFKDPIVNFELPIHLEKWFPRHVI
```

```
SEG   ............................................................
PRD   hhhhhccccccccccccccccccccccceeeeehhhhhhhcccceeeccccccccccccchhh

SEQ   KTKREDIRELILKLHLQQQKGSSS
SEG   ........................
PRD   hhhhhhhhhhhhhhhhhhhccccc
```

Prosite for DKFZphfbr2_16g18.3

PS00001	314->318	ASN_GLYCOSYLATION	PDOC00001
PS00001	365->369	ASN_GLYCOSYLATION	PDOC00001
PS00001	406->410	ASN_GLYCOSYLATION	PDOC00001
PS00001	440->444	ASN_GLYCOSYLATION	PDOC00001
PS00001	513->517	ASN_GLYCOSYLATION	PDOC00001
PS00001	600->604	ASN_GLYCOSYLATION	PDOC00001
PS00001	752->756	ASN_GLYCOSYLATION	PDOC00001
PS00001	759->763	ASN_GLYCOSYLATION	PDOC00001
PS00001	790->794	ASN_GLYCOSYLATION	PDOC00001
PS00001	830->834	ASN_GLYCOSYLATION	PDOC00001
PS00001	856->860	ASN_GLYCOSYLATION	PDOC00001
PS00001	922->926	ASN_GLYCOSYLATION	PDOC00001
PS00004	8->12	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	21->25	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	54->57	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00005	88->91	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00005	162->165	PKC_PHOSPHO_SITE	PDOC00005
PS00005	172->175	PKC_PHOSPHO_SITE	PDOC00005
PS00005	233->236	PKC_PHOSPHO_SITE	PDOC00005
PS00005	236->239	PKC_PHOSPHO_SITE	PDOC00005
PS00005	260->263	PKC_PHOSPHO_SITE	PDOC00005
PS00005	291->294	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	515->518	PKC_PHOSPHO_SITE	PDOC00005
PS00005	562->565	PKC_PHOSPHO_SITE	PDOC00005
PS00005	602->605	PKC_PHOSPHO_SITE	PDOC00005
PS00005	747->750	PKC_PHOSPHO_SITE	PDOC00005
PS00005	874->877	PKC_PHOSPHO_SITE	PDOC00005
PS00005	879->882	PKC_PHOSPHO_SITE	PDOC00005
PS00005	901->904	PKC_PHOSPHO_SITE	PDOC00005
PS00005	962->965	PKC_PHOSPHO_SITE	PDOC00005
PS00006	11->15	CK2_PHOSPHO_SITE	PDOC00006
PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00006	91->95	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006
PS00006	167->171	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	225->229	CK2_PHOSPHO_SITE	PDOC00006
PS00006	251->255	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	295->299	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	341->345	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	396->400	CK2_PHOSPHO_SITE	PDOC00006
PS00006	402->406	CK2_PHOSPHO_SITE	PDOC00006
PS00006	408->412	CK2_PHOSPHO_SITE	PDOC00006
PS00006	488->492	CK2_PHOSPHO_SITE	PDOC00006
PS00006	509->513	CK2_PHOSPHO_SITE	PDOC00006
PS00006	536->540	CK2_PHOSPHO_SITE	PDOC00006
PS00006	562->566	CK2_PHOSPHO_SITE	PDOC00006
PS00006	602->606	CK2_PHOSPHO_SITE	PDOC00006
PS00006	638->642	CK2_PHOSPHO_SITE	PDOC00006
PS00006	664->668	CK2_PHOSPHO_SITE	PDOC00006
PS00006	697->701	CK2_PHOSPHO_SITE	PDOC00006
PS00006	747->751	CK2_PHOSPHO_SITE	PDOC00006
PS00006	826->830	CK2_PHOSPHO_SITE	PDOC00006
PS00006	846->850	CK2_PHOSPHO_SITE	PDOC00006
PS00006	962->966	CK2_PHOSPHO_SITE	PDOC00006
PS00007	216->223	TYR_PHOSPHO_SITE	PDOC00007
PS00008	84->90	MYRISTYL	PDOC00008
PS00008	106->112	MYRISTYL	PDOC00008
PS00008	141->147	MYRISTYL	PDOC00008
PS00008	161->167	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	468->474	MYRISTYL	PDOC00008

PS00008	505->511	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00008	693->699	MYRISTYL	PDOC00008
PS00009	6->10	AMIDATION	PDOC00009
PS00009	18->22	AMIDATION	PDOC00009
PS00009	109->113	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_16g18.3)

DKFZphfbr2_16i12
----

group: transmembrane protein

DKFZphfbr2_16i12 encodes a novel 185 amino acid protein, with strong similarity to PUT2 protein of Fugu rubripes.

The novel protein contains 1 transmembrane region.
PUT 2 is a Fugu rupies protein similar to the neural cell adhesion molecule L1 (L1-CAM) a mitosis-specific chromosome segregation protein (SMC1) and the calcium channel alpha-1 subunit homolog (CCA1).
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

strong similarity to Fugu rubripes PUT2

complete cDNA, complete cds, EST hits,
TRANSMEMBRANE 1

Sequenced by LMU

Locus: /map="873.3/875.1 cR from top of Chr1 linkage group"

Insert length: 1552 bp
Poly A stretch at pos. 1528, polyadenylation signal at pos. 1506

```
   1 GGGGGGGGAC AACTGGGTCT TTTGCGGCTG CAGCGGGCTT GTAGGCGTCC
  51 GGCTTTGCTG GCCCAGCAAG CCTGATAAGC ATGAAGCTCT TATCTTTGGT
 101 GGCTGTGGTC GGGTGTTTGC TGGTGCCCCC AGCTGAAGCC AACAAGAGTT
 151 CTGAAGATAT CCGGTGCAAA TGCATCTGTC CACCTTATAG AAACATCAGT
 201 GGGCACATTT ACAACCAGAA TGTATCCCAG AAGGACTGTT GTAGCAACTG
 251 CCTGCACGTG GTGGAGCCCA TGCCAGTGCC TGGCCATGAC GTGGAGGCCT
 301 ACTGCCTGCT GTGCGAGTGC AGGTACGAGG AGCGCAGCAC CACCACCATC
 351 AAGGTCATCA TTGTCATCTA CCTGTCCGTG GTGGGTGCCC TGTTGCTCTA
 401 CATGGCCTTC CTGATGCTGG TGGACCCTCT GATCCGAAAG CCGGATGCAT
 451 ACACTGAGCA ACTGCACAAT GAGGAGGAGA ATGAGGATGC TCGCTCTATG
 501 GCAGCAGCTG CTGCATCCCT CGGGGGACCC CGAGCAAACA CAGTCCTGGA
 551 GCGTGTGGAA GGTGCCCAGC AGCGGTGGAA GCTGCAGGTG CAGGAGCAGC
 601 GGAAGACAGT CTTCGATCGG CACAAGATGC TCAGCTAGAT GGGCTGGTGT
 651 GGTTGGGTCA AGGCCCCAAC ACCATGGCTG CCAGCTTCCA GGCTGGACAA
 701 AGCAGGGGGC TACTTCTCCC TTCCCTCGGT TCCAGTCTTC CCTTTAAAAG
 751 CCTGTGGCAT TTTTCCTCCT TCTCCCTAAC TTTAGAAATG TTGTACTTGG
 801 CTATTTTGAT TAGGGAAGAG GGATGTGGTC TCTGATCTCT GTTGTCTTCT
 851 TGGGTCTTTG GGGTTGAAGG GAGGGGGAAG GCAGGCCAGA AGGGAATGGA
 901 GACATTCGAG GCGGCCTCAG GAGTGGATGC GATCTGTCTC TCCTGGCTCC
 951 ACTCTTGCCG CCTTCCAGCT CTGAGTCTTG GGAATGTTGT TACCCTTGGA
1001 AGATAAAGCT GGGTCTTCAG GAACTCAGTG TTTGGGAGGA AAGCATGGCC
1051 CAGCATTCAG CATGTGTTCC TTTCTGCAGT GGTTCTTATC ACCACCTCCC
1101 TCCCAGCCCC AGCGCCTCAG CCCCAGCCCC AGCTCCAGCC CTGAGGACAG
1151 CTCTGATGGG AGAGCTGGGC CCCCTGAGCC CACTGGGTCT TCAGGGTGCA
1201 CTGGAAGCTG GTGTTCGCTG TCCCCTGTGC ACTTCTCGCA CTGGGGCATG
1251 GAGTGCCCAT GCATACTCTG CTGCCGGTCC CCTCACCTGC ACTTGAGGGG
1301 TCTGGGCAGT CCCTCCTCTC CCCAGTGTCC ACAGTCACTG AGCCAGACGG
1351 TCGGTTGGAA CATGAGACTC GAGGCTGAGC GTGGATCTGA ACACCACAGC
1401 CCCTGTACTT GGGTTGCCTC TTGTCCCTGA ACTTCGTTGT ACCAGTGCAT
1451 GGAGAGAAAA TTTTGTCCTC TTGTCTTAGA GTTGTGTGTA AATCAAGGAA
1501 GCCATCATTA AATTGTTTTA TTTCTCTCAA AAAAAAAAAA AAAAAAAATA
1551 TC
```

BLAST Results

Entry HS808349 from database EMBL:
human STS WI-11986.
Score = 1716, P = 5.7e-73, identities = 364/378

Entry HS487355 from database EMBL:
human STS WI-13088.
Score = 1358, P = 1.3e-56, identities = 274/277

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 81 bp to 635 bp; peptide length: 185
Category: similarity to unknown protein

```
  1 MKLLSLVAVV GCLLVPPAEA NKSSEDIRCK CICPPYRNIS GHIYNQNVSQ
 51 KDCCSNCLHV VEPMPVPGHD VEAYCLLCEC RYEERSTTTI KVIIVIYLSV
101 VGALLLYMAF LMLVDPLIRK PDAYTEQLHN EEENEDARSM AAAAASLGGP
151 RANTVLERVE GAQQRWKLQV QEQRKTVFDR HKMLS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_16i12, frame 3

TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete cds; putative protein 1 (PUT1) gene, partial cds; mitosis-specific chromosome segregation protein SMC1 homolog (SMC1) gene, complete cds; and calcium channel alpha-1 subunit homolog (CCA1) and putative protein 2 (PUT2) genes, partial cds, complete sequence., N = 1, Score = 655, P = 2.8e-64

TREMBL:CER12C12_5 gene: "R12C12.6"; Caenorhabditis elegans cosmid R12C12., N = 1, Score = 225, P = 1e-18

>TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete cds; putative protein 1 (PUT1) gene, partial cds; mitosis-specific chromosome segregation protein SMC1 homolog (SMC1) gene, complete cds; and calcium channel alpha-1 subunit homolog (CCA1) and putative protein 2 (PUT2) genes, partial cds, complete sequence.
Length = 187

HSPs:

Score = 655 (98.3 bits), Expect = 2.8e-64, P = 2.8e-64
Identities = 124/163 (76%), Positives = 140/163 (85%)

```
Query:    22 KSSEDIRCKCICPPYRNISGHIYNQNVSQKDCCSNCLHVVEPMPVPGHDVEAYCLLCECR 81
             KS +D+RCKCICPPYRNISGHIYN+N +QKDC  NCLHVV+PMPVPG+DVEAYCLLCEC+
Sbjct:    31 KSFDDVRCKCICPPYRNISGHIYNRNFTQKDC--NCLHVVDPMPVPGNDVEAYCLLCECK 88

Query:    82 YEERSTTTIKVIIVIYLSVVGALLLYMAFLMLVDPLIRKPDAYTEQLHNEEENEDARSMA 141
             YEERST TI+V I+I+LSVVGALLLYM FL+LVDPLIRKPD   + LHNEE++ED +
Sbjct:    89 YEERSTNTIRVTIIIFLSVVGALLLYMLFLLLVDPLIRKPDPLAQTLHNEEDSEDIQPQM 148

Query:   142 AAAASLGGP-RANTVLERVEGAQQRWKLQVQEQRKTVFDRHKML 184
             +     G P R NTVLERVEGAQQRWK QVQEQRKTVFDRHKML
Sbjct:   149 S-----GDPARGNTVLERVEGAQQRWKKQVQEQRKTVFDRHKML 187
```

Pedant information for DKFZphfbr2_16i12, frame 3

Report for DKFZphfbr2_16i12.3

[LENGTH] 185
[MW] 20764.29
[pI] 6.21
[HOMOL] TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete cds; putative protein 1 (PUT1) gene, partial cds; mitosis-specific chromosome segregation protein SMC1 homolog (SMC1) gene, complete cds; and calcium channel alpha-1 subunit homolog (CCA1) and putative protein 2 (PUT2) genes, partial cds, complete sequence. 3e-68
[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 4
[PROSITE] PKC_PHOSPHO_SITE 2
[PROSITE] ASN_GLYCOSYLATION 3
[KW] SIGNAL_PEPTIDE 21

```
[KW]          TRANSMEMBRANE  1
[KW]          LOW_COMPLEXITY      2.70 %

SEQ     MKLLSLVAVVGCLLVPPAEANKSSEDIRCKCICPPYRNISGHIYNQNVSQKDCCSNCLHV
SEG     ............................................................
PRD     ccceeeeeeeeccccccccccccccceeeeeecccccccccceeeccccccccccccceeee
MEM     ............................................................

SEQ     VEPMPVPGHDVEAYCLLCECRYEERSTTTIKVIIVIYLSVVGALLLYMAFLMLVDPLIRK
SEG     ............................................................
PRD     eeccccccccccchhhhhhhhhhhhcccccceeeeeeehhhhhhhhhhhhhhhhhhhhcccc
MEM     ..........................MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ     PDAYTEQLHNEEENEDARSMAAAAASLGGPRANTVLERVEGAQQRWKLQVQEQRKTVFDR
SEG     ...................xxxxx....................................
PRD     ccchhhhhhhhhccccchhhhhhhhhhccccccchhhhhhhchhhhhhhhhhhhhhhhhh
MEM     ............................................................

SEQ     HKMLS
SEG     .....
PRD     hhccc
MEM     .....
```

Prosite for DKFZphfbr2_16i12.3

PS00001	21->25	ASN_GLYCOSYLATION	PDOC00001
PS00001	38->42	ASN_GLYCOSYLATION	PDOC00001
PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00006	23->27	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	148->154	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16i12.3)

DKFZphfbr2_16k22
----

group: brain derived

DKFZphfbr2_16k22 encodes a novel 108 amino acid protein with very weak similarity to thioredoxin of Bacillus subtilis.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to thioredoxin

complete cDNA, complete cds, genomic DNA?
no EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 2088 bp
Poly A stretch at pos. 2065, no polyadenylation signal found

```
   1 AAAAGGAAGA AGGAAATAAG GATATTTCAA GGGTTACCAA AGTCGAGGAA
  51 AACTATTTTA AGAAGAAATC TGAATTATTT GTGCACATAG GTTGTAATAA
 101 TAGCATCTTG CATTAAATGG TGTTTTCTAG CTTACAAAGT GGATTCATAT
 151 ACACTATTGT AACTGACTCT CTACAAACTT GCAAGGTTAG CAAGACAAAT
 201 GGTATTTTAA GATAACAAAC TGAGACTCAA AAAAGGCAAG TAACTCGTTC
 251 TACTTCCCAA AGCCAGAAAG TGGCAAAATA GAAAATGGAT CCTGAATCTC
 301 CAACACCATG CAAACTAAGA GAGGGAATCC TCTGTAGAGG GAATGGAAGT
 351 AAAAAGGCAC AAGTGGTGAT GTCACCTTCT GAACAGAGAT GGAACTTTTC
 401 TTCCTCTGAG AAAAAAGAGA AAAGATAGTT TTAAGTGGCA AAAGAACATG
 451 AAGCAATGTG AGGTGAAGAA ACAGAAAAGA CTATGGATGG AATTCCTAGA
 501 TGTGAGATAC ACAAAGTTCC ATTTCAAAGA GAAATATCTA TAGATAGGCA
 551 TAAAGTTACA CACCTGAACT ACCAACTCTG AACCAGTAAC TCAAGAGATA
 601 TTTTGTGTGT CCCACAAGCC ATATGGCTCT GGGGACAAAT TATCTGAAAG
 651 TGCCCAATAA GAAAAATATT TGAGGAAGGG GAGTTGGTGA GTGAATGAAT
 701 TAAAGGACAT CAGAAAGATA CATTGACTGT TCTCCTTCCC AGGAAACAAA
 751 GTGGCTAAGT CAAAACAACG GGCAGCTGTG GGATAGCAAA GAAAAAAAAA
 801 CTTCCAGGCC CAGGTTCTAG TGAAAGCTAC TATGGAAGTT AGCCACTCAA
 851 CTTTAGAACC AGAGGCTTCT TTTCCTCCTC CCTTCTTATC TTTTCTAGTT
 901 TATAGCAAAT TTATATTGAG CCACTTATTC TTTCTGAATG CTAGTTCCCC
 951 TTTAGCATTT CTTTTTCTTC ATTCCCTTTG GACTGGCCCA ATGCTTTGGC
1001 CCCTTATCAA AGCATTTTCT AAGAAACAGT CTGACAGCTC TAATTTGCAT
1051 CTGGTTATGC AAGATGTGGT TAAGAACATG GACTCTGGAG GTAAATACAC
1101 CTTGATTCCA ATTCATTCTC TCATTTATTC ATTCAGCAAA TATTTAGTGA
1151 ACATCTAACA TGTGCTAGGC ACTGTTCTAG TTGCTGAGGA TACAGCTTCA
1201 AACAAAATAA GGTCTCTGCA AGGATGCCTT CTCTTACCAC TCCTATTCAG
1251 CGTAGTATTG GAAGTCCTGG CCAGGGCAAT CAGGCAAGAA AAAGAAATCA
1301 AGGTCATCCA AATAGGAAGA GAGGAAGTCA AACTATCCCT GTTTACAGAC
1351 AACATGATCC TACATCTAGA AAAAAACCCA TTGTCTTAGC CCAAAAGCTT
1401 CTTAGGCTGA TAAACAACTT CAGCAAAGTC TTAGGATACA AAATCCATGT
1451 GCAAAAAACA CTAGCATTCT TATACACCAA CAACAGTCAA GCCGAGATCC
1501 AAATCAGGAA CAAACTCCTA TTCACAATTG CCACAAAAAC AATAGAACAG
1551 GAAAACAGCT AACTAGGAAG GTGAAAGATC TCTACAAGGA GAACTACAAA
1601 CCACTGCTCA CAGAAATCAG AGATGACACA TATAAATGGA AAAACATTCC
1651 ATGATCATGG ATAGGAAGAA TGAATATTAC TGAAATGGCT ATACTGTCCA
1701 AAGCAATTTA TAGATTCAAT GCTATTCCTA GTAAACTACC ATTGAGATTT
1751 TTTACAGAAC TAGAAAAAAA AAAAACTATT TTAAGGCTGG GCGCAGTGGC
1801 TCTCACCTGT AATCCCAGCA CTTTGGGAGG CCGAGATGGG TGGATCACGA
1851 GGTCAGGAGA TGGAAAACAT CCTGGCTAAC ATGGTGAAAC CCCGTCTCTA
1901 CTAAAAATAC AAAAAATTAG CCAGGCGTGG TGGTGGGCGC CTGTAATCCC
1951 AGCTGCTCGG GAGGCTGAGG CAGGATAATG GTGTGAACCC GGGAGGCAGA
2001 GCTTGCAGTG AGCTGAGATT GCACCACTGC ACTCCAGCCT GAGGGACAGA
2051 GTGAGACTCC ATCTCAAAAA AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 832 bp to 1155 bp; peptide length: 108
Category: putative protein

```
    1 MEVSHSTLEP EASFPPPFLS FLVYSKFILS HLFFLNASSP LAFLFLHSLW
   51 TGPMLWPLIK AFSKKQSDSS NLHLVMQDVV KNMDSGGKYT LIPIHSLIYS
  101 FSKYLVNI
```

BLASTP hits

Entry B37192 from database PIR:
thioredoxin - Bacillus subtilis Score = 71 (25.0 bits), Expect = 0.040,
P = 0.039
Identities = 16/49 (32%), Positives = 30/49 (61%)

Alert BLASTP hits for DKFZphfbr2_16k22, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_16k22, frame 1

Report for DKFZphfbr2_16k22.1

[LENGTH]	108	
[MW]	12281.47	
[pI]	8.06	
[PROSITE]	MYRISTYL	1
[PROSITE]	CAMP_PHOSPHO_SITE	1
[PROSITE]	CK2_PHOSPHO_SITE	1
[PROSITE]	PKC_PHOSPHO_SITE	1
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	Alpha_Beta	

```
SEQ  MEVSHSTLEPEASFPPPFLSFLVYSKFILSHLFFLNASSPLAFLFLHSLWTGPMLWPLIK
PRD  cccccccccccccccccccchhhhhhhhhhhhhhhhccccchhhhhhhhcccccccchhhhh

SEQ  AFSKKQSDSSNLHLVMQDVVKNMDSGGKYTLIPIHSLIYSFSKYLVNI
PRD  hhhcccccccceeehhhhhhcccccccceeeeecccceeeecccccccc
```

Prosite for DKFZphfbr2_16k22.1

PS00001	36->40	ASN_GLYCOSYLATION	PDOC00001
PS00004	64->68	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00006	6->10	CK2_PHOSPHO_SITE	PDOC00006
PS00008	86->92	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16k22.1)

DKFZphfbr2_16112
----

group: transmembrane protein

DKFZphfbr2_16112 encodes a novel 267 amino acid protein with similarity to gallus gallus putative transmembrane protein E3-16

The novel protein contains one putative transmembrane domain. In chicken, E3-16 is expressed specifically in the inner ear.

No informative BLAST results; no predictive prosite, pfam or SCOP motife

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neurons involved in perception of hearing.

similarity to gallus putative transmembrane protein E3-16

complete cDNA, complete cds, EST hits
potental start at Bp 73 matchs kozak consensus PyCCataG
TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 2042 bp
Poly A stretch at pos. 2024, polyadenylation signal at pos. 2003

```
   1 GGGGGCGGCG GAGGCAGAGA CCGAGGCTGC ACCGGCAGAG GCTGCGGGGC
  51 GGACGCGCGG GCCGGCGCAG CCATGGTGAA GATTAGCTTC CAGCCCGCCG
 101 TGGCTGGCAT CAAGGGCGAC AAGGCTGACA AGGCGTCGGC GTCGGCCCCT
 151 GCGCCGGCCT CGGCCACCGA GATCCTGCTG ACGCCGGCTA GGGAGGAGCA
 201 GCCCCCACAA CATCGATCCA AGAGGGGGGG CTCAGTGGGC GGCGTGTGCT
 251 ACCTGTCGAT GGGCATGGTC GTGCTGCTCA TGGGCCTCGT GTTCGCCTCT
 301 GTCTACATCT ACAGATACTT CTTCCTTGCG CAGCTGGCCC GAGATAACTT
 351 CTTCCGCTGT GGTGTGCTGT ATGAGGACTC CCTGTCCTCC CAGGTCCGGA
 401 CTCAGATGGA GCTGGAAGAG GATGTGAAAA TCTACCTCGA CGAGAACTAC
 451 GAGCGCATCA ACGTGCCTGT GCCCCAGTTT GGCGGCGGTG ACCCTGCAGA
 501 CATCATCCAT GACTTCCAGC GGGGTCTGAC TGCGTACCAT GATATCTCCC
 551 TGGACAAGTG CTATGTCATC GAACTCAACA CCACCATTGT GCTGCCCCCT
 601 CGCAACTTCT GGGAGCTCCT CATGAACGTG AAGAGGGGGA CCTACCTGCC
 651 GCAGACGTAC ATCATCCAGG AGGAGATGGT GGTCACGGAG CATGTCAGTG
 701 ACAAGGAGGC CCTGGGGTCC TTCATCTACC ACCTGTGCAA CGGGAAAGAC
 751 ACCTACCGGC TCCGGCGCCG GGCAACGCGG AGGCGGATCA ACAAGCGTGG
 801 GGCCAAGAAC TGCAATGCCA TCCGCCACTT CGAGAACACC TTCGTGGTGG
 851 AGACGCTCAT CTGCGGGGTG GTGTGAGGCC CTCCTCCCCC AGAACCCCCT
 901 GCCGTGTTCC TCTTTTCTTC TTTCCGGCTG CTCTCTGGCC CTCCTCCTTC
 951 CCCCTGCTTA GCTTGTACTT TGGACGCGTT TCTATAGAGG TGACATGTCT
1001 CTCCATTCCT CTCCAACCCT GCCCACCTCC CTGTACCAGA GCTGTGATCT
1051 CTCGGTGGGG GGCCCATCTC TGCTGACCTG GGTGTGGCGG AGGGAGAGGC
1101 GATGCTGCAA AGTGTTTTCT GTGTCCCACT GTCTTGAAGC TGGGCCTGCC
1151 AAAGCCTGGG CCCACAGCTG CACCGGCAGC CCAAGGGGAA GGACCGGTTG
1201 GGGGAGCCGG GCATGTGAGG CCCTGGGCAA GGGGATGGGG CTGTGGGGGC
1251 GGGGCGGCAT GGGCTTCAGA AGTATCTGCA CAATTAGAAA AGTCCTCAGA
1301 AGCTTTTTCT TGGAGGGTAC ACTTTCTTCA CTGTCCCTAT TCCTAGACCT
1351 GGGGCTTGAG CTGAGGATGG GACGATGTGC CCAGGGAGGG ACCCACCAGA
1401 GCACAAGAGA AGGTGGCTAC CTGGGGGTGT CCCAGGGACT CTGTCAGTGC
1451 CTTCAGCCCA CCAGCAGGAG CTTGGAGTTT GGGGAGTGGG GATGAGTCCG
1501 TCAAGCACAA CTGTTCTCTG AGTGGAACCA AAGAAGCAAG GAGCTAGGAC
1551 CCCCAGTCCT GCCCCCCAGG AGCACAAGCA GGGTCCCCTC AGTCAAGGCA
1601 GTGGGATGGG CGGCTGAGGA ACGGGGCAGG CAAGGTCACT GCTCAGTCAC
1651 GTCCACGGGG GACGAGCCGT GGGTTCTGCT GAGTAGGTGG AGCTCATTGC
1701 TTTCTCCAAG CTTGGAACTG TTTTGAAAGA TAACACAGAG GGAAAGGGAG
1751 AGCCACCTGG TACTTGTCCA CCCTGCCTCC TCTGTTCTGA AATTCCATCC
1801 CCCTCAGCTT AGGGGAATGC ACCTTTTTCC CTTTCCTTCT CACTTTTGCA
1851 TGTTTTTACT GATCATTCGA TATGCTAACC GTTCTCAGCC CTGAGCCTTG
1901 GAGAGGAGGG CTGTAACGCC TTCAGTCAGT CTCTGGGGAT GAAACTCTTA
1951 AATGCTTTGT ATATTTTCTC AATTAGATCT CTTTTCAGAA GTGTCTATAG
2001 AACAATAAAA ATCTTTTACT TCTGAAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

96325063:
Isolation of markers for chondro-osteogenic differentiation using cDNA library subtraction. Molecular cloning and characterization of a gene belonging to a novel multigene family of integral membrane proteins.

Peptide information for frame 1

ORF from 73 bp to 873 bp; peptide length: 267
Category: similarity to known protein

```
  1 MVKISFQPAV AGIKGDKADK ASASAPAPAS ATEILLTPAR EEQPPQHRSK
 51 RGGSVGGVCY LSMGMVVLLM GLVFASVYIY RYFFLAQLAR DNFFRCGVLY
101 EDSLSSQVRT QMELEEDVKI YLDENYERIN VPVPQFGGGD PADIIHDFQR
151 GLTAYHDISL DKCYVIELNT TIVLPPRNFW ELLMNVKRGT YLPQTYIIQE
201 EMVVTEHVSD KEALGSFIYH LCNGKDTYRL RRRATRRRIN KRGAKNCNAI
251 RHFENTFVVE TLICGVV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_16112, frame 1

SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16)., N = 1, Score = 573, P = 1.4e-55

SWISSNEW:ITMB_MOUSE INTEGRAL MEMBRANE PROTEIN 2B (E25B PROTEIN)., N = 1, Score = 559, P = 4.2e-54

SWISSNEW:ITMA_HUMAN INTEGRAL MEMBRANE PROTEIN 2A (E25 PROTEIN)., N = 1, Score = 452, P = 9.1e-43

>SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).
Length = 262

HSPs:

Score = 573 (86.0 bits), Expect = 1.4e-55, P = 1.4e-55
Identities = 118/264 (44%), Positives = 175/264 (66%)

```
Query:    1 MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGGSVGGVCY 60
            MVK+SF  A+A     + A+K    ++        ++L+ P  + + P+    G     C+
Sbjct:    1 MVKVSFNSALA--HKEAANKEEENS-------QVLILPP-DAKEPEDVVVPAGHKRAWCW 50

Query:   61 -LSMGMVVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLY-EDSLS-----SQVRTQM- 112
             +  G+  +L G++    Y+Y+YF    Q       + CG+ Y ED LS      +Q+++
Sbjct:   51 CMCFGLAFMLAGVILGGAYLYKYFAFQQ---GGVYFCGIKYIEDGLSLPESGAQLKSARY 107

Query:  113 -ELEEDVKIYLDENYERINVPVPQFGGGDPADIIHDFQRGLTAYHDISLDKCYVIELNTT 171
             +E++++I  +E+ E  I+VPVP+F    DPADI+HDF R LTAY D+SLDKCYVI LNT+
Sbjct:  108 HTIEQNIQILEEEDVEFISVPVPEFADSDPADIVHDFHRRLTAYLDLSLDKCYVIPLNTS 167

Query:  172 IVLPPRNFWELLMNVKRGTYLPQTYIIQEEMVVTEHVSDKEALGSFIYHLCNGKDTYRLR 231
            +V+PP+NF ELL+N+K GTYLPQ+Y+I E+M+VT+ + + + LG FIY LC GK+TY+L+
Sbjct:  168 VVMPPKNFLELLINIKAGTYLPQSYLIHEQMIVTDRIENVDQLGFFIYRLCRGKETYKLQ 227

Query:  232 RRATRRRINKRGAKNCNAIRHFENTFVVETLIC 264
            R+    + I KR A NC  IRHFEN F +ETLIC
Sbjct:  228 RKEAMKGIQKREAVNCRKIRHFENRFAMETLIC 260
```

Pedant information for DKFZphfbr2_16112, frame 1

Report for DKFZphfbr2_16112.1

[LENGTH] 267
[MW] 30223.94

```
[pI]            8.16
[HOMOL]         SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).
1e-49
[PROSITE]       PRENYLATION     1
[PROSITE]       MYRISTYL        5
[PROSITE]       CAMP_PHOSPHO_SITE      2
[PROSITE]       CK2_PHOSPHO_SITE       3
[PROSITE]       TYR_PHOSPHO_SITE       1
[PROSITE]       PKC_PHOSPHO_SITE       4
[PROSITE]       ASN_GLYCOSYLATION      1
[KW]            TRANSMEMBRANE   1
[KW]            LOW_COMPLEXITY    15.36 %

SEQ   MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGGSVGGVCY
SEG   ..............xxxxxxxxxxxxxxx...............................
PRD   ccccccccchhhhhhhhhhhhhhhhhhcccccceeeccccccccccccccccccccccchh
MEM   ....................................................MMMMMMMM

SEQ   LSMGMVVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLYEDSLSSQVRTQMELEEDVKI
SEG   ..xxxxxxxxxxx...............................................
PRD   hhhhhhhhhhhhhhhhhhcchhhhhhhhhhccceeeeeecccccccchhhhhhhhhhhh
MEM   MMMMMMMMMMMMMMMMMMM.........................................

SEQ   YLDENYERINVPVPQFGGGDPADIIHDFQRGLTAYHDISLDKCYVIELNTTIVLPPRNFW
SEG   ............................................................
PRD   hhccccceeeecccccccccccccchhhhhhhhhhhhhhhhccccceeeeeccceeeccccchhh
MEM   ............................................................

SEQ   ELLMNVKRGTYLPQTYIIQEEMVVTEHVSDKEALGSFIYHLCNGKDTYRLRRRATRRRIN
SEG   .................................................xxxxxxxxxxx
PRD   hhhhhhcccccccceeeeehhhhhhhcccccchhhhhheeeccccchhhhhhhhhhhhhh
MEM   ............................................................

SEQ   KRGAKNCNAIRHFENTFVVETLICGVV
SEG   xx.........................
PRD   hhhhccceeeeccccchhhhhheeeccc
MEM   ...........................
```

Prosite for DKFZphfbr2_16112.1

PS00001	169->173	ASN_GLYCOSYLATION	PDOC00001
PS00004	187->191	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	227->230	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00007	119->127	TYR_PHOSPHO_SITE	PDOC00007
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	53->59	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	243->249	MYRISTYL	PDOC00008
PS00294	264->268	PRENYLATION	PDOC00266

(No Pfam data available for DKFZphfbr2_16112.1)

DKFZphfbr2_22f21
----

group: brain derived

DKFZphfbr2_22f21 encodes a novel 567 amino acid protein with weak similarity to C. elegans cosmide C18C4.5

No informative BLAST results; no predictive prosite, pfam or SCOP motife

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to C.elegans C18C4.5

EST HSAA6531/HSAA5273/ defines splice variant, or unspliced cDNA additional ~180 Bp at position 250

Sequenced by AGOWA

Locus: /map="311.4 cR from top of Chrl4 linkage group"

Insert length: 1910 bp
Poly A stretch at pos. 1887, polyadenylation signal at pos. 1867

```
   1 TGGGCCCTTA GCAACGGCCT GGCGACGGTT TCCTTGCTGC TGCAGCCCCC
  51 GTCGGCTCCT CTTTTCCAGT CCTCCACTGC CGGGGCTGGG CCCGGCCGCG
 101 GGAAGGACCG AAGGGGATAC AGCGTGTCCC TGCGGCGGCT GCAAGAGGAC
 151 TAAGCATGGA TGGCAGCCGG AGAGTCAGAG CAACCTCTGT CCTTCCCAGA
 201 TATGGTCCAC CGTGCCTATT TAAAGGACAC TTGAGCACCA AAAGTAATGC
 251 TGCAGTAGAC TGCTCGGTTC CAGTAAGCAT GAGTACCAGC ATAAAGTATG
 301 CAGACCAACA ACGAAGAGAG AAACTCAAAA AGGAATTAGC ACAATGTGAA
 351 AAAGAGTTCA AATTAACTAA AACTGCAATG CGAGCCAATT ATAAAAATAA
 401 TTCCAAGTCA CTTTTTAATA CCTTACAAGA GCCCTCAGGC GAACCGCAAA
 451 TTGAGGATGA CATGTTAAAA GAAGAAATGA ATGGATTTTC ATCCTTTGCA
 501 AGGTCACTAG TACCCTCTTC AGAGAGACTA CACCTAAGTC TACATAAATC
 551 CAGTAAAGTC ATCACAAATG GTCCTGAGAA GAACTCCAGT TCCTCCCCGT
 601 CCAGTGTGGA TTATGCAGCC TCCGGGCCCC GGAAACTGAG CTCTGGAGCC
 651 CTGTATGGCA GAAGGCCCAG AAGCACATTC CCAAATTCCC ACCGGTTTCA
 701 GTTAGTCATT TCGAAAGCAC CCAGTGGGGA TCTTTTGGAT AAACATTCTG
 751 AACTCTTTTC TAACAAACAA TTGCCATTCA CTCCTCGCAC TTTAAAAACA
 801 GAAGCAAAAT CTTTCCTGTC ACAGTATCGC TATTATACAC CTGCCAAAAG
 851 AAAAAAGGAT TTTACAGATC AACGGATAGA AGCTGAAACC CAGACTGAAT
 901 TAAGCTTTAA ATCTGAGTTG GGGACAGCTG AGACTAAAAA CATGACAGAT
 951 TCAGAAATGA ACATAAAGCA GGCATCTAAT TGTGTGACAT ATGATGCCAA
1001 AGAAAAAATA GCTCCTTTAC CTTTAGAAGG GCATGACTCA ACATGGGATG
1051 AGATTAAGGA TGATGCTCTT CAGCATTCCT CACCAAGGGC AATGTGTCAG
1101 TATTCCCTGA AGCCCCCTTC AACTCGTAAA ATCTACTCTG ATGAAGAAGA
1151 ACTGTTGTAT CTGAGTTTCA TTGAAGATGT AACAGATGAA ATTTTGAAAC
1201 TTGGTTTATT TTCAAACAGG TTTTTAGAAC GACTGTTCGA GCGACATATA
1251 AAACAAAATA AACATTTGGA GGGGGAAAAA ATGCGCCACC TGCTGCATGT
1301 CCTGAAAGTA GACTTAGGCT GCACATCGGA GGAAAACTCG GTAAAGCAAA
1351 ATGATGTTGA TATGTTGAAT GTATTTGATT TTGAAAAGGC TGGGAATTCA
1401 GAACCAAATA AATTAAAAAA TGAAAGTGAA GTAACAATTC AGCAGGAACG
1451 TCAACAATAC CAAAAGGCTT TGGATATGTT ATTGTCGGCA CCAAAGGATG
1501 AGAACGAGAT ATTCCCTTCA CCAACTGAAT TTTTCATGCC TATTTATAAA
1551 TCAAAGCATT CAGAAGGGGT TATAATTCAA CAGGTGAATG ATGAAACAAA
1601 TCTTGAAACT TCAACTTTGG ATGAAAATCA TCCAAGTATT TCAGACAGTT
1651 TAACAGATCG GGAAACTTCT GTGAATGTCA TTGAAGGTGA TAGTGACCCT
1701 GAAAAGGTTG AGATTTCAAA TGGATTATGT GGTCTTAACA CATCACCCTC
1751 CCAATCTGTT CAGTTCTCCA GTGTCAAAGG CGACAATAAT CATGACATGG
1801 AGTTATCAAC TCTTAAAATC ATGGAAATGA GCATTGAGGA CTGCCCTTTG
1851 GATGTTTAAT CTTCATTAAT AAATACCTCA AATGGCCAGT AAAAAAAAAA
1901 AAAAAAAAAA
```

BLAST Results

Entry HS477360 from database EMBL:
human STS WI-14643.
Length = 418
Minus Strand HSPs:
Score = 1850 (277.6 bits), Expect = 2.5e-77, P = 2.5e-77
Identities = 392/405 (96%), Positives = 392/405 (96%), Strand = Minus / Plus

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 156 bp to 1856 bp; peptide length: 567
Category: similarity to unknown protein

```
  1 MDGSRRVRAT SVLPRYGPPC LFKGHLSTKS NAAVDCSVPV SMSTSIKYAD
 51 QQRREKLKKE LAQCEKEFKL TKTAMRANYK NNSKSLFNTL QEPSGEPQIE
101 DDMLKEEMNG FSSFARSLVP SSERLHLSLH KSSKVITNGP EKNSSSSPSS
151 VDYAASGPRK LSSGALYGRR PRSTFPNSHR FQLVISKAPS GDLLDKHSEL
201 FSNKQLPFTP RTLKTEAKSF LSQYRYYTPA KRKKDFTDQR IEAETQTELS
251 FKSELGTAET KNMTDSEMNI KQASNCVTYD AKEKIAPLPL EGHDSTWDEI
301 KDDALQHSSP RAMCQYSLKP PSTRKIYSDE EELLYLSFIE DVTDEILKLG
351 LFSNRFLERL FERHIKQNKH LEGEKMRHLL HVLKVDLGCT SEENSVKQND
401 VDMLNVFDFE KAGNSEPNKL KNESEVTIQQ ERQQYQKALD MLLSAPKDEN
451 EIFPSPTEFF MPIYKSKHSE GVIIQQVNDE TNLETSTLDE NHPSISDSLT
501 DRETSVNVIE GDSDPEKVEI SNGLCGLNTS PSQSVQFSSV KGDNNHDMEL
551 STLKIMEMSI EDCPLDV
```

BLASTP hits

Entry CEC18C4_3 from database TREMBL:
"C18C4.5"; Caenorhabditis elegans cosmid C18C4.
Length = 1091
Score = 98 (34.5 bits), Expect = 0.29, P = 0.25
Identities = 105/470 (22%), Positives = 192/470 (40%)

Alert BLASTP hits for DKFZphfbr2_22f21, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_22f21, frame 3

Report for DKFZphfbr2_22f21.3

[LENGTH]	567	
[MW]	64120.02	
[pI]	5.68	
[PROSITE]	AMIDATION	1
[PROSITE]	MYRISTYL	3
[PROSITE]	CAMP_PHOSPHO_SITE	1
[PROSITE]	CK2_PHOSPHO_SITE	16
[PROSITE]	PKC_PHOSPHO_SITE	18
[PROSITE]	ASN_GLYCOSYLATION	4
[KW]	All_Alpha	
[KW]	LOW_COMPLEXITY	1.23 %

```
SEQ  MDGSRRVRATSVLPRYGPPCLFKGHLSTKSNAAVDCSVPVSMSTSIKYADQQRREKLKKE
SEG  ............................................................
PRD  cccccceeeeeecccccccccccccccccccceeeecccccccchhhhhhhhhhhhhhhh

SEQ  LAQCEKEFKLTKTAMRANYKNNSKSLFNTLQEPSGEPQIEDDMLKEEMNGFSSFARSLVP
SEG  ............................................................
PRD  hhhhhhhhhhhhhhhhhhhhcccccceeecccccccchhhhhhhhhhhcccccceeecc

SEQ  SSERLHLSLHKSSKVITNGPEKNSSSSPSSVDYAASGPRKLSSGALYGRRPRSTFPNSHR
SEG  .......................xxxxxxx..............................
PRD  ccchhhhhhhhceeeeccccccccccccccccccccccccccccccccccccccccccccc

SEQ  FQLVISKAPSGDLLDKHSELFSNKQLPFTPRTLKTEAKSFLSQYRYYTPAKRKKDFTDQR
SEG  ............................................................
PRD  cceeeeeccccccccccccccccccccccccccchhhhhhhhhhhhcccccchhhhhhhh

SEQ  IEAETQTELSFKSELGTAETKNMTDSEMNIKQASNCVTYDAKEKIAPLPLEGHDSTWDEI
SEG  ............................................................
PRD  hhhhhhhhhhhhhhcccccccccchhhhhhhhccceeehhhhhhcccccccccccccccc
```

```
SEQ   KDDALQHSSPRAMCQYSLKPPSTRKIYSDEEELLYLSFIEDVTDEILKLGLFSNRFLERL
SEG   ............................................................
PRD   ccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhh

SEQ   FERHIKQNKHLEGEKMRHLLHVLKVDLGCTSEENSVKQNDVDMLNVFDFEKAGNSEPNKL
SEG   ............................................................
PRD   hhhhhhhhhhcccchhhhhhhhhcccccccccccccccccccceeeecccccccccccc

SEQ   KNESEVTIQQERQQYQKALDMLLSAPKDENEIFPSPTEFFMPIYKSKHSEGVIIQQVNDE
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccceeeeecccc

SEQ   TNLETSTLDENHPSISDSLTDRETSVNVIEGDSDPEKVEISNGLCGLNTSPSQSVQFSSV
SEG   ............................................................
PRD   ccccccccccccccccccccccceeecccccccccccceeeecccccccccccceeeeecc

SEQ   KGDNNHDMELSTLKIMEMSIEDCPLDV
SEG   ...........................
PRD   ccccccchhhhhhhhhhhhhccccccc
```

Prosite for DKFZphfbr2_22f21.3

PS00001	81->85	ASN_GLYCOSYLATION	PDOC00001
PS00001	143->147	ASN_GLYCOSYLATION	PDOC00001
PS00001	262->266	ASN_GLYCOSYLATION	PDOC00001
PS00001	422->426	ASN_GLYCOSYLATION	PDOC00001
PS00004	159->163	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	45->48	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	132->135	PKC_PHOSPHO_SITE	PDOC00005
PS00005	178->181	PKC_PHOSPHO_SITE	PDOC00005
PS00005	202->205	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	212->215	PKC_PHOSPHO_SITE	PDOC00005
PS00005	250->253	PKC_PHOSPHO_SITE	PDOC00005
PS00005	309->312	PKC_PHOSPHO_SITE	PDOC00005
PS00005	317->320	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	353->356	PKC_PHOSPHO_SITE	PDOC00005
PS00005	395->398	PKC_PHOSPHO_SITE	PDOC00005
PS00005	500->503	PKC_PHOSPHO_SITE	PDOC00005
PS00005	539->542	PKC_PHOSPHO_SITE	PDOC00005
PS00005	552->555	PKC_PHOSPHO_SITE	PDOC00005
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	149->153	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	264->268	CK2_PHOSPHO_SITE	PDOC00006
PS00006	295->299	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00006	337->341	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	481->485	CK2_PHOSPHO_SITE	PDOC00006
PS00006	486->490	CK2_PHOSPHO_SITE	PDOC00006
PS00006	494->498	CK2_PHOSPHO_SITE	PDOC00006
PS00006	498->502	CK2_PHOSPHO_SITE	PDOC00006
PS00006	500->504	CK2_PHOSPHO_SITE	PDOC00006
PS00006	513->517	CK2_PHOSPHO_SITE	PDOC00006
PS00006	559->563	CK2_PHOSPHO_SITE	PDOC00006
PS00008	164->170	MYRISTYL	PDOC00008
PS00008	256->262	MYRISTYL	PDOC00008
PS00008	350->356	MYRISTYL	PDOC00008
PS00009	167->171	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_22f21.3)

DKFZphfbr2_22h13
----

group: transmembrane protein

DKFZphfbr2_22h13 encodes a novel 520 amino acid protein, with similarity to Drosophila melanogaster EG:39E1.3.

The protein contains an ATP/GTP A Prosite pattern (P-loop). This loop interacts with one of the phosphate groups of a A or G nucleotide. It is found in numerous ATP- or GTP-binding proteins, such as ATP synthase alpha and beta subunits, Myosin heavy chains, Kinesin heavy chains and kinesin-like proteins, Dynamins and dynamin-like proteins, several kinases, DNA and RNA helicases, GTP-binding elongation factors and the Ras family of GTP-binding proteins. Additionally, the novel protein contains one putative transmembran domain.

No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

AC004780_1, differences to predicted genmodel

membrane regions: 1

AC004780_1, differences to predicted genmodel

complete cDNA, complete cds, EST hits
on genomic level encoded by AC004780,
differences to predicted genmodel!
TRANSMEMBRANE 1

Sequenced by AGOWA

Locus: unknown

Insert length: 2292 bp
Poly A stretch at pos. 2272, polyadenylation signal at pos. 2255

```
   1 GGGGGAGGGA ACTGATCTCA GCTCGGGCCC GCGTTACATC CTCCTCCTCT
  51 TCTTCCTTCG GCCCAGCTTT CCTTAGGGGC TGCAACCCGG ACGCCGAGGC
 101 CGGTTTCGGA GTGGGGAGTG CCCATTTTCT CTCCTTCCCA CGTTCCTGGC
 151 CCCCAGACGC CATTTGCAGG CGGGTGGCTT GGGTCAGCCT CCCCGCCCCC
 201 ACCCGACTCC CGTCACGGGA GAGCGCACAC CGCGCCCCGA GAACCAATCA
 251 GCAGCCGCGT TAGGTAACCA TGTCTGAGTC TGGACACAGT CAGCCTGGAC
 301 TCTATGGGAT AGAGCGGCGG CGACGGTGGA AGGAGCCTGG CTCTGGTGGC
 351 CCCCAGAATC TCTCTGGGCC TGGTGGTCGG GAGAGGGACT ACATTGCACC
 401 ATGGGAAAGA GAGAGAAGGG ATGCCAGCGA AGAGACAAGC ACTTCCGTCA
 451 TGCAGAAAAC CCCCATCATC CTCTCAAAAC CTCCAGCAGA GCGGTCAAAA
 501 CAGCCACCAC CTCCAACAGC CCCTGCTGCC CCGCCTGCTC CAGCCCCTCT
 551 GGAGAAGCCC ATCGTTCTCA TGAAGCCACG GGAGGAGGGG AAGGGGCCTG
 601 TGGCCGTGAC AGGTGCCTCT ACCCCTGAGG GCACCGCCCC ACCACCCCCT
 651 GCAGCCCCTG CGCCACCCAA GGGGGAGAAG GAGGGGCAGA GACCCACACA
 701 GCCTGTGTAC CAGATCCAGA ACCGGGGCAT GGGCACTGCC GCACCAGCAG
 751 CCATGGACCC TGTCGTGGGT CAGGCCAAAC TACTGCCCCC AGAGCGCATG
 801 AAGCACAGCA TCAAGTTGGT GGATGACCAG ATGAATTGGT GTGACAGTGC
 851 CATCGAGTAC CTGTTGGATC AGACTGATGT GTTGGTGGTT GGTGTCCTGG
 901 GCCTCCAGGG GACAGGCAAG TCCATGGTCA TGTCATTGTT GTCAGCCAAC
 951 ACTCCAGAGG AGGACCAGAG GACTTATGTT TTCCGGGCCC AGAGCGCTGA
1001 AATGAAGGAA CGAGGGGGCA ACCAGACCAG TGGCATCGAC TTCTTTATTA
1051 CCCAAGAACG GATTGTTTTC CTGGACACAC AGCCCATCCT GAGCCCTTCT
1101 ATCCTAGACC ATCTCATCAA TAATGACCGC AAACTGCCTC CAGAGTACAA
1151 CCTTCCCCAC ACTTACGTTG AAATGCAGTC ACTCCAGATT GCTGCCTTCC
1201 TTTTCACGGT CTGCCATGTG GTGATTGTTG TCCAGGACTG GTTCACAGAC
1251 CTCAGTCTCT ACAGGTTCCT GCAGACAGCA GAGATGGTGA AGCCCTCCAC
1301 CCCATCCCCC AGCCACGAGT CCAGCAGCTC ATCGGGCTCC GATGAAGGCA
1351 CCGAGTACTA CCCCCACCTA GTCTTCTTGC AGAACAAAGC TCGCCGAGAG
1401 GACTTCTGTC CTCGGAAGCT GCGGCAGATG CACCTGATGA TTGACCAGCT
1451 CATGGCCCAC TCCCACCTGC GTTACAAGGG AACTCTGTCC ATGTTACAAT
1501 GCAATGTCTT CCCGGGGCTT CCACCTGACT TCCTGGACTC TGAGGTCAAC
1551 TTATTCCTGG TACCCTTCAT GGACAGTGAA GCAGAGAGTG AAAACCCACC
1601 AAGAGCAGGA CCTGGTTCCA GCCCACTCTT CTCCCTGCTG CCTGGGTATC
1651 GTGGCCACCC CAGTTTCCAG TCCTTGGTGA GCAAGCTCCG GAGCCAAGTG
1701 ATGTCCATGG CCCGGCCACA GCTGTCACAC ACGATCCTCA CCGAGAAGAA
1751 CTGGTTCCAC TACGCTGCCC GGATCTGGGA TGGGGTGAGA AAGTCCTCTG
1801 CTCTGGCAGA GTACAGCCGC CTGCTGGCCT GAGGCCAAGG AGAGGAATGT
1851 CATGCAGGGG ACCTCCTGGG TCCGCAGTGT ACTGCGAGGG AGCACAGATG
1901 TCCATCCCCC GCTGGGGTGG AGAGCGGCAG CAGGCCTGAT GGATGAGGGA
1951 TCGTGGCTTC CCGGCCCAGA GACATGAGGT GTCCAGGGCC AGGCCCCCCA
```

```
2001 CCCTCAGTTG GGGCTGTTCC GGGGGTGACT GTGAGCGATC CCACCCCAAA
2051 CCTGAGATGG GGTAGCCCGT CCTGTGTCCT CCACAGGGAC AAGCAGTGGG
2101 AGGAGTCTGA ATGGTCACCA GGAAGCCCGG GCTCCATCTT GACCTCCTTT
2151 TTCAGGGACA GGAGCAACAG GCCCCTCTTC CCTGACTCTA AGCCCTTCCC
2201 TGTAAGGTGA GGCAGGGTCT GGAGAGCTCT TTATTGGAAC AGATCTGGTG
2251 GTTCAAATAA ACACAGTCAT GCAAAAAAAA AAAAAAAAAA AA
```

BLAST Results

Entry AC004780 from database EMBL:
Homo sapiens chromosome 19, cosmid F17127, complete sequence.
Score = 2616, P = 0.0e+00, identities = 524/525
15 exons Bp 8031-31789

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 270 bp to 1829 bp; peptide length: 520
Category: similarity to unknown protein
Prosite motifs: ATP_GTP_A (211-219)

```
  1 MSESGHSQPG LYGIERRRRW KEPGSGGPQN LSGPGGRERD YIAPWERERR
 51 DASEETSTSV MQKTPIILSK PPAERSKQPP PPTAPAAPPA PAPLEKPIVL
101 MKPREEGKGP VAVTGASTPE GTAPPPPAAP APPKGEKEGQ RPTQPVYQIQ
151 NRGMGTAAPA AMDPVVGQAK LLPPERMKHS IKLVDDQMNW CDSAIEYLLD
201 QTDVLVVGVL GLQGTGKSMV MSLLSANTPE EDQRTYVFRA QSAEMKERGG
251 NQTSGIDFFI TQERIVFLDT QPILSPSILD HLINNDRKLP PEYNLPHTYV
301 EMQSLQIAAF LFTVCHVVIV VQDWFTDLSL YRFLQTAEMV KPSTPSPSHE
351 SSSSSGSDEG TEYYPHLVFL QNKARREDFC PRKLRQMHLM IDQLMAHSHL
401 RYKGTLSMLQ CNVFPGLPPD FLDSEVNLFL VPFMDSEAES ENPPRAGPGS
451 SPLFSLLPGY RGHPSFQSLV SKLRSQVMSM ARPQLSHTIL TEKNWFHYAA
501 RIWDGVRKSS ALAEYSRLLA
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_22h13, frame 3

TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19, cosmid F17127, complete sequence., N = 2, Score = 1264, P = 1.3e-231

TREMBL:CEY54E2A_1 gene: "Y54E2A.2"; Caenorhabditis elegans cosmid Y54E2A, N = 2, Score = 219, P = 1.4e-15

>TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19, cosmid F17127, complete sequence.
Length = 528

HSPs:

Score = 1264 (189.6 bits), Expect = 1.3e-231, Sum P(2) = 1.3e-231
Identities = 254/302 (84%), Positives = 264/302 (87%)

```
Query:   46 ERERRDASEETSTSVMQKTPIILSKPPAERSKQPPPPTAPAAPPAPAPLEKPIVLMKPRE 105
            E+ER D+  + S   +Q+T    +    R  + P      +  A APLEKPIVLMKPRE
Sbjct:   39 EKER-DSDSDFSP--LQQTEGCQRRDKHFRHAENPHHPLKTSSRA-APLEKPIVLMKPRE 94

Query:  106 EGKGPVAVTGASTPEGTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAPAAMDPV 165
            EGKGPVAVTGASTPEGTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAPAAMDPV
Sbjct:   95 EGKGPVAVTGASTPEGTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAPAAMDPV 154

Query:  166 VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS 225
            VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS
Sbjct:  155 VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS 214
```

```
Query:   226 ANTPEEDQRTYVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINN 285
             ANTPEEDQRTYVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINN
Sbjct:   215 ANTPEEDQRTYVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINN 274

Query:   286 DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFTDLSLYRFLQTAEMVKPSTP 345
             DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFTDLSLYR          K ++
Sbjct:   275 DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFTDLSLYRLWDLGCKCKSNSH 334

Query:   346 SP 347
             SP
Sbjct:   335 SP 336

 Score = 993 (149.0 bits), Expect = 1.3e-231, Sum P(2) = 1.3e-231
 Identities = 189/189 (100%), Positives = 189/189 (100%)

Query:   332 RFLQTAEMVKPSTPSPSHESSSSSGSDEGTEYYPHLVFLQNKARREDFCPRKLRQMHLMI 391
             RFLQTAEMVKPSTPSPSHESSSSSGSDEGTEYYPHLVFLQNKARREDFCPRKLRQMHLMI
Sbjct:   340 RFLQTAEMVKPSTPSPSHESSSSSGSDEGTEYYPHLVFLQNKARREDFCPRKLRQMHLMI 399

Query:   392 DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFLVPFMDSEAESENPPRAGPGSS 451
             DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFLVPFMDSEAESENPPRAGPGSS
Sbjct:   400 DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFLVPFMDSEAESENPPRAGPGSS 459

Query:   452 PLFSLLPGYRGHPSFQSLVSKLRSQVMSMARPQLSHTILTEKNWFHYAARIWDGVRKSSA 511
             PLFSLLPGYRGHPSFQSLVSKLRSQVMSMARPQLSHTILTEKNWFHYAARIWDGVRKSSA
Sbjct:   460 PLFSLLPGYRGHPSFQSLVSKLRSQVMSMARPQLSHTILTEKNWFHYAARIWDGVRKSSA 519

Query:   512 LAEYSRLLA 520
             LAEYSRLLA
Sbjct:   520 LAEYSRLLA 528
```

Pedant information for DKFZphfbr2_22h13, frame 3

Report for DKFZphfbr2_22h13.3

```
[LENGTH]        520
[MW]            57650.81
[pI]            6.52
[HOMOL]         TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19, cosmid
F17127, complete sequence. 0.0
[PROSITE]       ATP_GTP_A       1
[PROSITE]       MYRISTYL        8
[PROSITE]       CAMP_PHOSPHO_SITE       1
[PROSITE]       CK2_PHOSPHO_SITE        8
[PROSITE]       GLYCOSAMINOGLYCAN       1
[PROSITE]       PKC_PHOSPHO_SITE        3
[PROSITE]       ASN_GLYCOSYLATION       2
[KW]            TRANSMEMBRANE  1
[KW]            LOW_COMPLEXITY    11.73 %

SEQ   MSESGHSQPGLYGIERRRRWKEPGSGGPQNLSGPGGRERDYIAPWERERRDASEETSTSV
SEG   ............................................................
PRD   cccccccccccccccccccccccccccccccccccccceeeeehhhhhhhhhcccccccee
MEM   ............................................................

SEQ   MQKTPIILSKPPAERSKQPPPPTAPAAPPAPAPLEKPIVLMKPREEGKGPVAVTGASTPE
SEG   ..................xxxxxxxxxxxxxx............................
PRD   eecccceeeccccccccccccccccccccccccccccceeeeecccccccccceeeeccccccc
MEM   ............................................................

SEQ   GTAPPPPAAPAPPKGEKEGQRPTQPVYQIQNRGMGTAAPAAMDPVVGQAKLLPPERMKHS
SEG   ..xxxxxxxxxxx...............................................
PRD   cccccccccccccccccccccccceeeeeeccccccccccccccceeecceeeccccchhhhh
MEM   ............................................................

SEQ   IKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLSANTPEEDQRTYVFRA
SEG   .................xxxxxxxxxxxxxxxxxxx........................
PRD   hhhhccccchhhhhhhhhhcccccceeeeeeecccccccchhhhhhhhccccchhhhhheeee
MEM   ............................................................

SEQ   QSAEMKERGGNQTSGIDFFITQERIVFLDTQPILSPSILDHLINNDRKLPPEYNLPHTYV
SEG   ............................................................
PRD   hhhhhhhcccccceeeeeeeecceeeeeecccccccccccccccccccccccccccchh
MEM   ............................................................

SEQ   EMQSLQIAAFLFTVCHVVIVVQDWFTDLSLYRFLQTAEMVKPSTPSPSHESSSSSGSDEG
SEG   ........................................xxxxxxxxxxxxxxxx...
```

```
PRD     hhhhhhhhhhhhhheeeeeeeccchhhhhhhhhhhhhhcccccccccccccccccccccc
MEM     MMMMMMMMMMMMMMMMMMMMMMM.....................................

SEQ     TEYYPHLVFLQNKARREDFCPRKLRQMHLMIDQLMAHSHLRYKGTLSMLQCNVFPGLPPD
SEG     ............................................................
PRD     cccccceeeehhhhhhcccccchhhhhhhhhhhhhhhhcccccccccccccccccccccc
MEM     ............................................................

SEQ     FLDSEVNLFLVPFMDSEAESENPPRAGPGSSPLFSLLPGYRGHPSFQSLVSKLRSQVMSM
SEG     ............................................................
PRD     chhhhhheeeeeccccccccccccccccccceeeccccccccchhhhhhhhhhhhhhhhhh
MEM     ............................................................

SEQ     ARPQLSHTILTEKNWFHYAARIWDGVRKSSALAEYSRLLA
SEG     ........................................
PRD     hhhhhhheeeccchhhhhhhhhhhcchhhhhhhhhhhccc
MEM     ........................................
```

Prosite for DKFZphfbr2_22h13.3

PS00001	30->34	ASN_GLYCOSYLATION	PDOC00001
PS00001	251->255	ASN_GLYCOSYLATION	PDOC00001
PS00002	32->36	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	507->511	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00005	215->218	PKC_PHOSPHO_SITE	PDOC00005
PS00005	491->494	PKC_PHOSPHO_SITE	PDOC00005
PS00006	117->121	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	254->258	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00006	298->302	CK2_PHOSPHO_SITE	PDOC00006
PS00006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS00006	436->440	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	211->217	MYRISTYL	PDOC00008
PS00008	214->220	MYRISTYL	PDOC00008
PS00008	249->255	MYRISTYL	PDOC00008
PS00008	356->362	MYRISTYL	PDOC00008
PS00008	505->511	MYRISTYL	PDOC00008
PS00017	211->219	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphfbr2_22h13.3)

DKFZphfbr2_22i4
----

group: brain derived

DKFZphfbr2_22i4.1 encodes a novel 228 amino acid protein with similarity to the N-terminus of human p52rIPK.

No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to Human P52rIPK N-terminus

complete cDNA, complete cds, few EST hits
function of P52rIPK, repressor of p58IPK protein kinase inhibitor
upstream regulator of interferon induced proteins

Sequenced by AGOWA

Locus: unknown

Insert length: 4748 bp
Poly A stretch at pos. 4726, polyadenylation signal at pos. 4709

```
   1 TGGGTCCGGT CCTAGGGTCA CACCCACCGC AGGGTCTGGC TTGGTACAGT
  51 TGGGTGCATG CAGAAGTAGG TGGAGCTGCT GTTGCAGCCT TGAGAGAGTT
 101 TTATTGTAAA ACTCTTGTAA TTTATAGTAA TCGGAGGGGA AAACACCTCT
 151 TCCTTTTAAT TGCTCTGAGG ACCGCTGCCA AAGAAACGCA GTAGATCCGC
 201 TCCCTCTTGG GGGCGGGGAG AAAGAACGGG TTGTGTCCGC CATGTTGGTG
 251 AAGTCAAGCG AAGGCGACTA GAGCTCCAGG AGGGCCAGTT CTGTGGGCTC
 301 TAGTCGGCCA TATTAATAAA GAGAAAGGGA AGGCTGACCG TCCTTCGCCT
 351 CCGCCCCCAC ATACACACCC CTTCTTCCCA CTCCGCTCTC ACGACTAAGC
 401 TCTCACGATT AAGGCACGCC TGCCTCGATT GTCCAGCCTC TGCCAGAAGA
 451 AAGCTTAGCA GCCAGCGCCT CAGTAGAGAC CTAAGGGCGC TGAATGAGTG
 501 GGAAAGGGAA ATGCCGACCA ATTGCGCTGC GGCGGGCTGT GCCACTACCT
 551 ACAACAAGCA CATTAACATC AGCTTCCACA GGTTTCCTTT GGATCCTAAA
 601 AGAAGAAAAG AATGGGTTCG CCTGGTTAGG CGCAAAAATT TTGTGCCAGG
 651 AAAACACACT TTTCTTTGTT CAAAGCACTT TGAAGCCTCC TGTTTTGACC
 701 TAACAGGACA AACTCGACGA CTTAAAATGG ATGCTGTTCC AACCATTTTT
 751 GATTTTTGTA CCCATATAAA GTCTATGAAA CTCAAGTCAA GGAATCTTTT
 801 GAAGAAAAAC AACAGTTGTT CTCCAGCTGG ACCATCTAAT TTAAAATCAA
 851 ACATTAGTAG TCAGCAAGTA CTACTTGAAC ACAGCTATGC CTTTAGGAAT
 901 CCTATGGAGG CAAAAAAGAG GATCATTAAA CTGGAAAAAG AAATAGCAAG
 951 CTTAAGAAGA AAAATGAAAA CTTGCCTACA AAAGGAACGC AGAGCAACTC
1001 GAAGATGGAT CAAAGCCACG TGTTTGGTAA AGAATTTAGA AGCAAATAGT
1051 GTATTACCTA AAGGTACATC AGAACACATG TTACCAACTG CCTTAAGCAG
1101 TCTTCCCTTG GAAGATTTTA AGATCCTTGA ACAAGATCAA CAAGATAAAA
1151 CACTGCTAAG TCTAAATCTA AAACAGACCA AGAGTACCTT CATTTAAATT
1201 TAGCTTGCAC AGAGCTTGAT GCCTATCCTT CATTCTTTTC AGAAGTAAAG
1251 ATAATTATGG CACTTATGCC AAAATTCATT ATTTAATAAA GTTTTACTTG
1301 AAGTAACATT ACTGAATTTG TGAAGACTTG ATTACAAAAG AATAAAAAAC
1351 TTCATATGGA AATTTTATTT GAAAATGAGT GGAAGTGCCT TACATTAGAA
1401 TTACGGACTT AAAAATTTTG CTAATAAATT GTGTGTTTGA AAGGTGTTTT
1451 TTGTTTTTGT CTTTTTAAAC TACTGTTAAA AGAACAGCTT ATGATAAGTA
1501 ATATGTTTAA CTTAGAGAAG AATTTTTTCC TGTACCAAAG TTGGCATATT
1551 GCATTCTAAA TAAGATGCTA AATAAGAGTT AACCAACATT CAACATGACC
1601 TTAAAACTGC TGGGTTTTGT ATTAATTAAA TTATAATTGG CACTGTGATT
1651 TGAAAAAATTT ATAGAAAAAA AGGTACAGGG CAAGTTTTTA AATTAAAACT
1701 TTCTATATTT TGTTTTACCA GTAAAAGTGA GCTTATCATG GCCTCTCTCA
1751 TAAGAATGAT TTTAAAATAG GTTGTAAAAT ATTTTGAAAA TATTTGAATG
1801 TGAAGTACCA TTGAGTCATC CAAACTAGGT AAGGCCTCAA GTACTTTAAA
1851 CTAGTAAAAT CTAGTAGCTG ATAATATTCA CCTAAGTAAG TGTTGTAAAA
1901 TAATTCAGAG TTCAGGACCT AGCTTAGATA AATGTATACT ACTCTTTTTC
1951 TCATAGTAAA AATCTTACAT TTCCAACTTC AAAATTGGTG CTTCCATATT
2001 TGTTGATAAC CAAAACTCCT AAGGTTTTTT GTTTTCTTTT TAACTACTTT
2051 CCAAATGCAT ACTATACCTC AGAAATAGTG TATCAATATA GTGGGCTTTT
2101 TTTTTCCTCT TCATAAACCC ACAGTAAAAT TTAATCACAG GAAACTACTT
2151 ATATCTTCAC ACTTTGTATT GATAACTTAA AATGGCATCA GTTTATCTTA
2201 GACATCAGCT TGCTTTTTAT CTCCTTTTTT AGTGAGTGAA ATAGAGCAAC
2251 TAGCATGCCT GTGTTCCCAG CTACTTGGGA GGCTAAGGTG GGAAGATCAA
2301 TTGAACCTAG GAGGTTGAGG CTATAGTGAG CTGTGATTGC ACGACTGCAC
2351 TCCAGCCTGG GCAATGGAGT GAGACTCCTG TCTCTAAAAC AGCAACAACA
2401 AAAATAAAGC AACCATAGTG CATAAGGGAA ATTAAATGTT CCCTATAGAA
2451 ATATGTGTAT GTCTGTGATA GTGGTATGCA AATGCTAATT ATTTTATAAA
2501 ATAAAAGTTC AGAACTATTC TTATCATTGC CACTTGAACA ATTAAAGGGT
2551 TTGCTTTATT TCACTAATGT TTAATAGGAA CCCTTTGCTT CAAACAGCTT
```

```
2601 TGTTGAAATC ATGTAAAAAT TTGTTAATAG AGAATCAAGT TATTTAACTC
2651 AACTTATTTA ATTCAAGCTT GTGATACTAA CATACAAAGG TAGCATAAAC
2701 CAAGTCATAA ATTGCTGTAA TCTTTCCTGT AGAGTAATAG CTACTTCATG
2751 ATTTTTTTAA AAATTTCATT TTTTTGCTAT TTAGGATTGC ATTTGCTTGG
2801 CTCCTAGTAA CAATTCTTTT ACAGTATTAG CACTCTCTTT ACTAAGGAAT
2851 GCCTCCCAAG GAAATGCAAA GGTAGGAAAA GTCTCTTAGA ATGCCCATGA
2901 GGTATTTAAA ACAGATATTT ATGAAAATCT TTTTGTGAAT GTTATAAATC
2951 TTGCTAGTTA TTTTATCTTT ATCTTAAGTA TTAGATGTAG TTCCTTGGAA
3001 TTGTCATTAC ATATTTATTT TTTTCTAGTG TGGTTTCAAA TAACTTTTTG
3051 CCAACATATA ATCATCATCA AACATTCACT GACCATATCT ATTTTATAAC
3101 TCAAAATAAG TTGGACAAAT AATCATTTTA ATAAAAACTA TTTTTTCCAA
3151 GTATAACCAC TGTCATGTGG TTCACCCTTC ACCCCAGATA CAAAACACTT
3201 ATTTGTGTAG CCCAGTTCCC ATCTACAGTA ATACCTTGAA ACCTTAATAA
3251 ATTTTAAAAA TCATAAAAAT AAAATATTGT AAAATACAAC AAATTTTGGA
3301 CAAGGTTACT TCATCTTCAT TCATTATTAC CTGACAGTAT TAAACTACTA
3351 CTCAATAATT TTAGAGTAAA CTTTTCTGTG TTTTCCCCGT GATTTTCATT
3401 GTGCTGTCCT GACAACATGC TCCAAACTCT TTGCATCAAA TTGTTTTATT
3451 AACATACATT TGTCTACCTT AAAACTAGCT TTATTCACAG AGAAAGACCT
3501 AAAAGGAGTC TATTAAAATG CTGCTTTCAG TTTGATAGTT TTTTTTTTAA
3551 TCACTCTGAC CATAAACTAA CTGAAATTAT AATGGATTTT TTTTCCTCTC
3601 CCGGTCACAA CACAGATCTT CTGTTCATTT GTTCTCTGTC TACTGGGCAC
3651 CAACCTCTAC AAAGAACCAG CCAAAGGCTA GGTACTTGAT ATAAAAAGGA
3701 ATATTACATT ATTTTCTGCC CTCAAGTTGC TCTATCTCCT GAAAGAAACA
3751 AGTAATATTT ATAATACAAT ATGATAAATG CTACAAAAGA AATAGCTGTA
3801 AAGTCCTTTG GTAAATGCTG TTGAATTGGA ATTCAGTAAG AACTATAAAC
3851 TGTAGACCTT TTTATAATCA AATGCTTTTG TCTTGAAACA AAACAGATTC
3901 CTCCTTATAT TGACTTAGCA AAGGAGGTAC AAGGACATTG GCATTTGACC
3951 TGAATTATGG TGTTTTATTG AATGAGCTAT AAGACAACAT TTTTACCCTT
4001 TAAAATGAAC ACTGAACAAA TGTGTTAATG GTATCTTTGT TAAAAGGAAA
4051 ACATAGCTAT AAATAAAATA CTACATCGAA ATCCAGCACT GGAGTTCATT
4101 TGAAATTTGA TATTTTGTGT AAAGTAACAA ACCTATTAAC ACAGATTTTT
4151 AAAATAACTC AGAATCGTAT AAAGCACTTT GGTACTTATT TGTTCTCTTT
4201 TCCCTTACAT TCTGTGTGGT AGGTGGTATT ATCTCTGATT TACACATGAA
4251 GACATCCTTG TTAATGCAAT TTATTTATTC ATTCGGGCAT TTACTGTGTG
4301 CCAACTTGCA AAAGGAATAG AAATGTCTGT GATCTAGATA GTTCTAGATT
4351 GAACATAGAT TTTCTGCCAA CAAATCCTCT CTGCTGTTCA CATTATCCTT
4401 TGTTTAACGT ATGAACCAGG TTACTAAAAT AGGATAAATC ATGTGTCTTA
4451 GAATATGAAA ATAGTAAGGT CTTTGAGGTC ACTTGATCTT CTCTAAGTAG
4501 ACTTTATAAT ATTGTGTTTT ATCTCATTTC TCAATATTAG AATACGGGTA
4551 GATTTTAATT TTGCTATAAT ATAGGAAATG GTTCATCTTT GTACCAAAAT
4601 ATTGCATTCT TCTGATATTT AGACAGTTGG AAACTTTCTA AAATTGAGGA
4651 TTTTGTAGTG TATACTAAAT AATTGCATAT TCAAAAAAAT GTATTCTGAG
4701 TATGGTGATA TTAAACATTT TTCCCCAAAA AAAAAAAAAA AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

98107671:
Regulation of interferon-induced protein kinase PKR: modulation of P58IPK inhibitory function by a novel protein, P52rIPK

Peptide information for frame 1

ORF from 511 bp to 1194 bp; peptide length: 228
Category: similarity to known protein

```
  1 MPTNCAAAGC ATTYNKHINI SFHRFPLDPK RRKEWVRLVR RKNFVPGKHT
 51 FLCSKHFEAS CFDLTGQTRR LKMDAVPTIF DFCTHIKSMK LKSRNLLKKN
101 NSCSPAGPSN LKSNISSQQV LLEHSYAFRN PMEAKKRIIK LEKEIASLRR
151 KMKTCLQKER RATRRWIKAT CLVKNLEANS VLPKGTSEHM LPTALSSLPL
201 EDFKILEQDQ QDKTLLSLNL KQTKSTFI
```

BLASTP hits

Entry AF007393_1 from database TREMBL:
product: "P52rIPK"; Homo sapiens P52rIPK mRNA, complete cds.
Score = 166, P = 2.5e-11, identities = 40/106, positives = 56/106

Alert BLASTP hits for DKFZphfbr2_22i4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_22i4, frame 1

Report for DKFZphfbr2_22i4.1

```
[LENGTH]     228
[MW]         26259.94
[pI]         10.17
[HOMOL]      TREMBL:AF007393_1 product: "P52rIPK"; Homo sapiens P52rIPK mRNA, complete cds.
1e-09
[PROSITE]    MYRISTYL     1
[PROSITE]    CAMP_PHOSPHO_SITE     1
[PROSITE]    CK2_PHOSPHO_SITE      2
[PROSITE]    PKC_PHOSPHO_SITE      4
[PROSITE]    ASN_GLYCOSYLATION     3
[KW]         All_Alpha
[KW]         LOW_COMPLEXITY     7.02 %
```

```
SEQ  MPTNCAAAGCATTYNKHINISFHRFPLDPKRRKEWVRLVRRKNFVPGKHTFLCSKHFEAS
SEG  ............................................................
PRD  cccccccccccccccccccceeeeccccchhhhhhhhhhhhccccceeehhhhhhhh

SEQ  CFDLTGQTRRLKMDAVPTIFDFCTHIKSMKLKSRNLLKKNNSCSPAGPSNLKSNISSQQV
SEG  ..........................xxxxxxxxxxxxxxxx..................
PRD  ccccccccccccccccccceeeecccchhhhhhhhhhcccccccccccccccccchhh

SEQ  LLEHSYAFRNPMEAKKRIIKLEKEIASLRRKMKTCLQKERRATRRWIKATCLVKNLEANS
SEG  ............................................................
PRD  hhhhccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeecccccc

SEQ  VLPKGTSEHMLPTALSSLPLEDFKILEQDQQDKTLLSLNLKQTKSTFI
SEG  ................................................
PRD  ccccccccccccccccccccchhhhhhcccccccccccccccccccc
```

Prosite for DKFZphfbr2_22i4.1

PS00001	19->23	ASN_GLYCOSYLATION	PDOC00001
PS00001	100->104	ASN_GLYCOSYLATION	PDOC00001
PS00001	114->118	ASN_GLYCOSYLATION	PDOC00001
PS00004	160->164	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	68->71	PKC_PHOSPHO_SITE	PDOC00005
PS00005	88->91	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	163->166	PKC_PHOSPHO_SITE	PDOC00005
PS00006	60->64	CK2_PHOSPHO_SITE	PDOC00006
PS00006	78->82	CK2_PHOSPHO_SITE	PDOC00006
PS00008	9->15	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_22i4.1)

DKFZphfbr2_22k3
----

group: brain derived

DKFZphfbr2_22k3 encodes a novel 538 amino acid protein with weak similarity to extensins.

No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to extensins

complete cDNA, complete cds, few EST hits
CpG Island in 5' UTR complete cDNA

Sequenced by AGOWA

Locus: unknown

Insert length: 2775 bp
Poly A stretch at pos. 2755, polyadenylation signal at pos. 2718

```
   1 GGGGCTGCCC GCGCGCTCCA CGGTGCAGAG CTCTAAGCGC GCGGGCTGGC
  51 AGGCTGCGGC GCGTCAAGGT CAGCCTGGAG CTGGGTGGCG GCCTGCCTGG
 101 GGGCGGGGGA CCCTACTGGA GGCCCGGGCT GGGGCCTCCC AGCGCCTCGG
 151 CCATATTGAA TAGCTTCGAC TGGACCGTCT TTGTCTGCGA AGTCCTGTCC
 201 CAAGTTCCAG CCGCGTCCCT GGGGCCTGGG GCAGGAAGAG TCGCTGGCAG
 251 CCCGCGCGCC CCAACTTGGA GCTGGGACAC CACGTTTCCA GCTTGGAGTG
 301 GGCCTTGAGC CTTGGGACTG ACCTCGCCCC CGGCTCACGT AGGCATCCTG
 351 GAAATTGATT CCCCCAAGTC CTTGGTGGGG GAGCCGGACT TGGTCAAGAC
 401 TGTACTTGTT GCAGGCGAAG AGATTGGAGG CGTTTGGCTC GTCCCTGGCT
 451 AGGGAGGTGA GACTCTCCGG TCAGCGTTGC TGGAACTCCC CCCATCCAGT
 501 CCCTCCCTCA AGACTAAGGG CTACAGTAGT TTGTTGGGGC TCATTGCCCC
 551 CTCACCCCAG ATATCACCCT GGAGATCTTA AAGACTCTCG AGAAAAGCCA
 601 CGTGGGGGGC TGGTTCCCCT GGGGCTTCCT GCCGTCCCCC GACTGCCTCA
 651 TTCTTTGGAG CGTCCCCGAT GTCTGCAAAG ATGTGGATTT GGACGTCCTC
 701 GTGGAAGCCC TAAAGCCCGT GGGGACATTT AAGAAGATCG GCAAGGTGTT
 751 CCGCAAGGAG GAGGACTCCA CGGTGGGGAT GCTGCAGATC GGGGAGGACG
 801 TCGACTATTT GCTCATCCCC CGGGAGGTCA GGCTGGCTGG GGGCGTCTGG
 851 AGAGTCATCT CTAAGCCCGC CACCAAGGAA GCAGAATTTC GGGAGCGGCT
 901 GACCCAGTTC CTGGAAGAAG AGGGCCGCAC CCTGGAGGAC GTGGCCCGCA
 951 TCATGGAGAA GAGCACCCCG CACCCGCCCC AGCCCCCCAA AAAGCCCAAG
1001 GAGCCCCGAG TGAGGAGGAG AGTGCAGCAG ATGGTGACTC CTCCGCCCCG
1051 GCTGGTCGTG GGCACGTACG ACAGCAGCAA CGCCAGCGAC AGCGAGTTCA
1101 GCGACTTCGA GACCTCCAGA GACAAGAGCC GCCAGGGCCC GCGGCGGGGC
1151 AAGAAGGTGC GCAAAATGCC CGTCAGCTAC CTGGGCAGCA AGTTCCTGGG
1201 AAGCGACCTG GAGAGTGAGG ATGATGAGGA ACTGGTCGAG GCCTTCCTCC
1251 GGCGACAGGA GAAGCAGCCC AGCGCGCCGC CTGCCCGCCG CCGCGTCAAC
1301 CTGCCAGTGC CCATGTTTGA GGACAACCTG GGGCCTCAGC TGTCCAAAGC
1351 GGACAGGTGG CGGGAGTATG TCAGCCAGGT GTCCTGGGGG AAGCTGAAGC
1401 GGAGGGTGAA GGGTTGGGCG CCGAGGGCGG GCCCCGGGGT GGGCGAGGCC
1451 CGGCTGGCCT CCACCGCAGT GGAGAGCGCA GGGGTATCAT CGGCGCCAGA
1501 GGGCACCAGC CCGGGGGATC GCTTGGGAAA CGCGGGAGAT GTTTGTGTGC
1551 CCCAGGCTTC CCCTAGGCGA TGGAGGCCCA AGATCAACTG GGCCTCCTTT
1601 CGGCGCCGCA GGAAGGAGCA GACAGCACCC ACAGGTCAGG GGGCAGACAT
1651 CGAGGCTGAT CAGGGGGGAG AGGCTGCAGA TAGTCAAAGG GAAGAGGCCA
1701 TAGCTGACCA GCGGGAAGGG GCTGCAGGTA ATCAGAGGGC TGGGGCCCCA
1751 GCTGACCAGG GGGCAGAGGC TGCAGATAAT CAGAGGGAAG AGGCTGCAGA
1801 TAATCAGAGG GCAGGGGCCC CAGCTGAGGA GGGGGCAGAG GCTGCAGATA
1851 ACCAGAGGGA AGAGGCTGCA GATAATCAGA GGGCAGAGGC CCCAGCTGAC
1901 CAGAGGTCAC AGGGCACAGA TAACCACAGG GAAGAGGCTG CAGATAATCA
1951 GAGGGCGGAG GCCCCAGCTG ACCAGGGGTC AGAGGTTACA GATAATCAAA
2001 GGGAAGAGGC CGTACATGAC CAGAGGGAAA GGGCCCCAGC TGTCCAGGGT
2051 GCAGATAATC AGAGGGCACA GGCCCGGGCT GGCCAGAGGG CAGAGGCTGC
2101 ACATAATCAG AGGGCAGGGG CCCCAGGTAT CCAGGAAGCT GAAGTCTCAG
2151 CTGCCCAAGG GACCACAGGA ACAGCTCCAG GAGCCAGGGC CCGGAAACAG
2201 GTCAAGACAG TGAGGTTCCA GACCCCTGGA CGCTTTTCGT GGTTTTGCAA
2251 GCGCCGGAGA GCCTTCTGGC ACACTCCCCG GTTGCCAACC CTGCCCAAGA
2301 GAGTCCCCAG GGCAGGAGAG GTCAGGAACC TCAGGGTGCT GAGGGCCGAG
2351 GCCAGAGCAG AAGCTGAGCA GGGAGAGCAA GAAGACCAGC TGTGAGGTGA
2401 GGGCTAGAGA CAGCCCACGG GCCCTCCCTC CAAGTGTGGG AGGGAGAGAT
2451 GCTCTGCCTC TGAACTTCAA AGTGGAGGTG GAGTGCTGGC CACGTCTCCA
2501 CCTAACAACC CTCTTTATTC TCTTGTTAAA GTTTTGTTCA TGCTTTGATT
2551 TTTTTTTAAA TTTTTTAGAG ACAGGGTCTC ACTCTGTTGC CCAGGCTGGA
2601 GTGCAGTGGC ATGATCATAA CTCACTGCAG CCTCAAACTT CTGGCCTCAA
2651 GTGATCCTCC TGCCTCGGCC TCCCAAAATG CTGGGATTAC AGATGTGAGC
```

```
2701 CACCACACAC ACCATCTGAT TAAAAAAAAA AAATACTGAT TCCCTGTAGC
2751 AACCCAAAAA AAAAAAAAAA AAAAA
```

BLAST Results

Entry HS164A7F from database EMBL:
H.sapiens CpG island DNA genomic MseI fragment, clone 164a7, forward read cpg164a7.ft1a .
Score = 740, P = 3.0e-25, identities = 150/151

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 779 bp to 2392 bp; peptide length: 538
Category: similarity to known protein

```
  1 MLQIGEDVDY LLIPREVRLA GGVWRVISKP ATKEAEFRER LTQFLEEEGR
 51 TLEDVARIME KSTPHPPQPP KKPKEPRVRR RVQQMVTPPP RLVVGTYDSS
101 NASDSEFSDF ETSRDKSRQG PRRGKKVRKM PVSYLGSKFL GSDLESEDDE
151 ELVEAFLRRQ EKQPSAPPAR RRVNLPVPMF EDNLGPQLSK ADRWREYVSQ
201 VSWGKLKRRV KGWAPRAGPG VGEARLASTA VESAGVSSAP EGTSPGDRLG
251 NAGDVCVPQA SPRRWRPKIN WASFRRRRKE QTAPTGQGAD IEADQGGEAA
301 DSQREEAIAD QREGAAGNQR AGAPADQGAE AADNQREEAA DNQRAGAPAE
351 EGAEAADNQR EEAADNQRAE APADQRSQGT DNHREEAADN QRAEAPADQG
401 SEVTDNQREE AVHDQRERAP AVQGADNQRA QARAGQRAEA AHNQRAGAPG
451 IQEAEVSAAQ GTTGTAPGAR ARKQVKTVRF QTPGRFSWFC KRRRAFWHTP
501 RLPTLPKRVP RAGEVRNLRV LRAEARAEAE QGEQEDQL
```

BLASTP hits

Entry RNU67136_1 from database TREMBL:
"A-kinase anchoring protein AKAP150"; Rattus norvegicus
A-kinase anchoring protein AKAP150 mRNA, complete cds. Rattus norvegicus (Norway rat)
Length = 714
Score = 182 (64.1 bits), Expect = 1.2e-10, P = 1.2e-10
Identities = 73/257 (28%), Positives = 104/257 (40%)

Alert BLASTP hits for DKFZphfbr2_22k3, frame 2

TREMBL:PFSANTY_1 product: "S-antigen"; Plasmodium falciparum KF1916
S-antigen gene, complete cds., N = 1, Score = 178, P = 3.7e-11

>TREMBL:PFSANTY_1 product: "S-antigen"; Plasmodium falciparum KF1916
S-antigen gene, complete cds.
Length = 285

HSPs:

Score = 178 (26.7 bits), Expect = 3.7e-11, P = 3.7e-11
Identities = 60/217 (27%), Positives = 97/217 (44%).

```
Query:   269 INWASFRRRRKEQTAPTGQGA-DIEADQGGEAADSQRE-EAIADQ---REGAAGNQRAGA 323
             +N  + +  + E      G+G  D E       E +D+  E E I  Q     E A  N+ AG+
Sbjct:    47 LNGKNGKGNKYEDLQEEGEGENDDEEHSNSEESDNDEENEIIVGQDGSNEKAGSNEEAGS 106

Query:   324 PADQGAEAADNQREEAADNQRAGAPAEEGA--EAADNQR----EEAADNQRAEAPADQRS 377
                  G+       E+A  N++AG+  E G+  EA  N+      EEA  N++A  +       S
Sbjct:   107 NEKAGSNEEAGSNEKAGSNEKAGSNEEAGSNEEAGSNEEAGSNEEAGSNEKAGSNEKAGS 166

Query:   378 QGTDNHREEAADNQRAEAPADQGSEVTDNQREEAVHDQRERAPAVQGADNQRAQAR--AG 435
                      EEA  N++A  +  + GS          E+A  +++  +       G+ N++A  +    AG
Sbjct:   167 NEKAGSNEEAGSNEKAGSNEEAGSNEKAGSNEKAGSNEKAGSNEEAGS-NEKAGSNEEAG 225

Query:   436 QRAEAAHNQRAGA---PGIQEAEVSAAQGTTGTA-PGA 469
```

```
              EA  N+ AG+    G  E   +  +G  GT  PG+
Sbjct:   226 SNEEAGSNEEAGSNEEAGSNEGSEAGTEGPKGTGGPGS 263

 Score = 173 (26.0 bits), Expect = 1.5e-10, P = 1.5e-10
 Identities = 51/190 (26%), Positives = 83/190 (43%)

Query:   279 KEQTAPTGQ-GADIEADQGGEAADSQREEAIADQREGAAGNQRAGAPADQGAEAADNQRE 337
             +E     GQ G++ +A     EA  +++     A   E A  N++AG+    G+     E
Sbjct:    83 EENEIIVGQDGSNEKAGSNEEAGSNEK----AGSNEEAGSNEKAGSNEKAGSNEEAGSNE 138

Query:   338 EAADNQRAGAPAEEGAEAADNQREEAADNQRAEAPADQRSQGTDNHREEAADNQRAEAPA 397
             EA  N+ AG+  E G+          E+A  N++A +   +  S        EEA  N++A +
Sbjct:  .139 EAGSNEEAGSNEEAGSNEKAGSNEKAGSNEKAGSNEEAGSNEKAGSNEEAGSNEKAGSNE 198

Query:   398 DQGSEVTDNQREEAVHDQRERAPAVQGADNQRAQARAGQRAEAAHNQRAGAPGIQEAEVS 457
               GS          EEA  +++  +     G++ +      AG   EA  N+ AG+     EA
Sbjct:   199 KAGSNEKAGSNEEAGSNEKAGSNEEAGSNEE-----AGSNEEAGSNEEAGSNEGSEAGTE 253

Query:   458 AAQGTTGTAPG 468
               +GT G   G
Sbjct:   254 GPKGTGGPGSG 264

 Score = 147 (22.1 bits), Expect = 1.6e-07, P = 1.6e-07
 Identities = 40/168 (23%), Positives = 70/168 (41%)

Query:   288 GADIEADQGGEAADSQR--EEAIADQREGAAGNQRAGAPADQGAEAADNQREEAADNQRA 345
             G++ EA     +A  +++       A   E A  N+ AG+  + G+          E+A  N++A
Sbjct:   111 GSNEEAGSNEKAGSNEKAGSNEEAGSNEEAGSNEEAGSNEEAGSNEKAGSNEKAGSNEKA 170

Query:   346 GAPAEEGAEAADNQREEAADNQRAEAPADQRSQGTDNHREEAADNQRAEAPADQGSEVTD 405
             G+  E G+          EEA  N++A +        S          EEA  N++A +  +  GS
Sbjct:   171 GSNEEAGSNEKAGSNEEAGSNEKAGSNEKAGSNEKAGSNEEAGSNEKAGSNEEAGSNEEA 230

Query:   406 NQREEAVHDQR--ERAPAVQGADNQRAQARAGQRAEAAHNQRAGAPGI 451
                EEA  ++          +  G +  +           G   E  +HN++          I
Sbjct:   231 GSNEEAGSNEEAGSNEGSEAGTEGPKGTGGPGSGGEHSHNKKKSKKSI 278

 Score = 101 (15.2 bits), Expect = 2.5e-02, P = 2.4e-02
 Identities = 26/100 (26%), Positives = 47/100 (47%)

Query:   281 QTAPTGQGADIEADQGGEAADSQREEAIADQREGAAGNQRAGAPADQGAEAADNQREEAA 340
             + A + + A     + G          EEA ++++ G+  N++AG+      G+          E+A
Sbjct:   162 EKAGSNEKAGSNEEAGSNEKAGSNEEAGSNEKAGS--NEKAGSNEKAGSNEEAGSNEKAG 219

Query:   341 DNQRAGAPAEEGAEAADNQREEAADNQRAEAPADQRSQGT 380
              N+ AG+  E G+          EEA  N+ +EA   +  +GT
Sbjct:   220 SNEEAGSNEEAGSNEEAGSNEEAGSNEGSEA-GTEGPKGT 258
```

Pedant information for DKFZphfbr2_22k3, frame 2

Report for DKFZphfbr2_22k3.2

```
[LENGTH]      538
[MW]          59402.19
[pI]          8.72
[HOMOL]       TREMBL:AF037364_1 gene: "MA1"; product: "paraneoplastic neuronal antigen MA1";
Homo sapiens paraneoplastic neuronal antigen MA1 (MA1) mRNA, complete cds. 4e-10
[PROSITE]     AMIDATION          1
[PROSITE]     MYRISTYL          12
[PROSITE]     CK2_PHOSPHO_SITE        11
[PROSITE]     PKC_PHOSPHO_SITE         6
[PROSITE]     ASN_GLYCOSYLATION        1
[KW]          All_Alpha
[KW]          LOW_COMPLEXITY     18.03 %

SEQ     MLQIGEDVDYLLIPREVRLAGGVWRVISKPATKEAEFRERLTQFLEEEGRTLEDVARIME
SEG     ............................................................
PRD     cccccccccccccccccccccceeeeeeeccccchhhhhhhhhhhhhhhcccchhhhhhhhh

SEQ     KSTPHPPQPPKKPKEPRVRRRVQQMVTPPPRLVVGTYDSSNASDSEFSDFETSRDKSRQG
SEG     .....xxxxxxxxxxxxxxxxxx.....................................
PRD     hccccccccccccccccchhhhhhhhhcccccceeeeecccccccccccccccccccccccc

SEQ     PRRGKKVRKMPVSYLGSKFLGSDLESEDDEELVEAFLRRQEKQPSAPPARRRVNLPVPMF
SEG     ....................xxxxxxxxxxx.............................
PRD     ccccccccccceeecccccccccccccchhhhhhhhhhhhhhhcccccchhhhhccccccc
```

```
SEQ  EDNLGPQLSKADRWREYVSQVSWGKLKRRVKGWAPRAGPGVGEARLASTAVESAGVSSAP
SEG  ............................................................
PRD  cccccccchhhhhhhhhheeeeccchhhhhhccccccccccchhhhhhhhhhhccccccc

SEQ  EGTSPGDRLGNAGDVCVPQASPRRWRPKINWASFRRRRKEQTAPTGQGADIEADQGGEAA
SEG  ............................................................
PRD  cccccccccccccceeeecccccccccccccchhhhhhhhhhhhcccccchhhhhccchhh

SEQ  DSQREEAIADQREGAAGNQRAGAPADQGAEAADNQREEAADNQRAGAPAEEGAEAADNQR
SEG  ....................xxxxxxxxxxxxx...........xxxxxxxxxxxx....
PRD  hhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhcccccchhhhhhhhhh

SEQ  EEAADNQRAEAPADQRSQGTDNHREEAADNQRAEAPADQGSEVTDNQREEAVHDQRERAP
SEG  ............................................................
PRD  hhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  AVQGADNQRAQARAGQRAEAAHNQRAGAPGIQEAEVSAAQGTTGTAPGARARKQVKTVRF
SEG  .......xxxxxxxxxxxxxx.................xxxxxxxxxxxxxx........
PRD  hhccchhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhccccccccchhhhhhhhhhh

SEQ  QTPGRFSWFCKRRRAFWHTPRLPTLPKRVPRAGEVRNLRVLRAEARAEAEQGEQEDQL
SEG  .........................................xxxxxxxxxxxxxx...
PRD  cccccceeehhhhhhcccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhccc
```

Prosite for DKFZphfbr2_22k3.2

PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005
PS00005	273->276	PKC_PHOSPHO_SITE	PDOC00005
PS00005	302->305	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	499->502	PKC_PHOSPHO_SITE	PDOC00005
PS00006	51->55	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	112->116	CK2_PHOSPHO_SITE	PDOC00006
PS00006	142->146	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	189->193	CK2_PHOSPHO_SITE	PDOC00006
PS00006	229->233	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	302->306	CK2_PHOSPHO_SITE	PDOC00006
PS00008	95->101	MYRISTYL	PDOC00008
PS00008	220->226	MYRISTYL	PDOC00008
PS00008	242->248	MYRISTYL	PDOC00008
PS00008	296->302	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	317->323	MYRISTYL	PDOC00008
PS00008	328->334	MYRISTYL	PDOC00008
PS00008	352->358	MYRISTYL	PDOC00008
PS00008	400->406	MYRISTYL	PDOC00008
PS00008	450->456	MYRISTYL	PDOC00008
PS00008	461->467	MYRISTYL	PDOC00008
PS00008	464->470	MYRISTYL	PDOC00008
PS00009	123->127	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_22k3.2)

DKFZphfbr2_22k8
----

group: brain derived

DKFZphfbr2_22k8 encodes a novel 172 amino acid protein without similarity to known proteins.

No informative BLAST results; no predictive prosite, pfam or SCOP motife

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: /map="7"

Insert length: 2789 bp
Poly A stretch at pos. 2769, polyadenylation signal at pos. 2756

```
   1 GGGGGAGCCA TGAGGCGCCA GCCTGCGAAG GTGGCGGCGC TGCTGCTCGG
  51 GCTGCTCTTG GAGTGCACAG AAGCCAAAAA GCATTGCTGG TATTTCGAAG
 101 GACTCTATCC AACCTATTAT ATATGCCGCT CCTACGAGGA CTGCTGTGGC
 151 TCCAGGTGCT GTGTGCGGGC CCTCTCCATA CAGAGGCTGT GGTACTTCTG
 201 GTTCCTTCTG ATGATGGGCG TGCTTTTCTG CTGCGGAGCC GGCTTCTTCA
 251 TCCGGAGGCG CATGTACCCC CCGCCGCTGA TCGAGGAGCC AGCCTTCAAT
 301 GTGTCCTACA CCAGGCAGCC CCCAAATCCC GGCCCAGGAG CCCAGCAGCC
 351 GGGGCCGCCC TATTACACTG ACCCAGGAGG ACCGGGGATG AACCCTGTCG
 401 GGAATTCCAC GGCAATGGCT TTCCAGGTCC CACCCAACTC ACCCCAGGGG
 451 AGTGTGGCCT GCCCGCCCCC TCCAGCCTAC TGCAACACGC CTCCGCCCCC
 501 GTACGAACAG GTAGTGAAGG CCAAGTAGTG GGGTGCCCAC GTGCAAGAGG
 551 AGAGACAGGA GAGGGCCTTT CCCTGGCCTT TCTGTCTTCG TTGATGTTCA
 601 CTTCCAGGAA CGGTCTCGTG GGCTGCTAAG GGCAGTTCCT CTGATATCCT
 651 CACAGCAAGC ACAGCTCTCT TTCAGGCTTT CCATGGAGTA CAATATATGA
 701 ACTCACACTT TGTCTCCTCT GTTGCTTCTG TTTCTGACGC AGTCTGTGCT
 751 CTCACATGGT AGTGTGGTGA CAGTCCCCGA GGGCTGACGT CCTTACGGTG
 801 GCGTGACCAG ATCTACAGGA GAGAGACTGA GAGGAAGAAG GCAGTGCTGG
 851 AGGTGCAGGT GGCATGTAGA GGGGCCAGGC CGAGCATCCC AGGCAAGCAT
 901 CCTTCTGCCC GGGTATTAAT AGGAAGCCCC ATGCCGGGCG GCTCAGCCGA
 951 TGAAGCAGCA GCCGACTGAG CTGAGCCCAG CAGGTCATCT GCTCCAGCCT
1001 GTCCTCTCGT CAGCCTTCCT CTTCCAGAAG CTGTTGGAGA GACATTCAGG
1051 AGAGAGCAAG CCCCTTGTCA TGTTTCTGTC TCTGTTCATA TCCTAAAGAT
1101 AGACTTCTCC TGCACCGCCA GGGAAGGATA GCACGTGCAG CTCTCACCGC
1151 AGGATGGGGC CTAGAATCAG GCTTGCCTTG GAGGCCTGAC AGTGATCTGA
1201 CATCCACTAA GCAAATTTAT TTAAATTCAT GGGAAATCAC TTCCTGCCCC
1251 AAACTGAGAC ATTGCATTTT GTGAGCTCTT GGTCTGATTT GGAGAAAGGA
1301 CTGTTACCCA TTTTTTTGGT GTGTTTATGG AAGTGCATGT AGAGCGTCCT
1351 GCCCTTTGAA ATCAGACTGG GTGTGTGTCT TCCCTGGACA TCACTGCCTC
1401 TCCAGGGCAT TCTCAGGCCC GGGGGTCTCC TTCCCTCAGG CAGCTCCAGT
1451 GGTGGGTTCT GAAGGGTGCT TTCAAAACGG GGCACATCTG GCCGGGAAGT
1501 CACATGGACT CTTCCAGGGA GAGAGACCAG CTGAGGCGTC TCTCTCTGAG
1551 GTTGTGTTGG GTCTAAGCGG GTGTGTGCTG GGCTCCAAGG AGGAGGAGCT
1601 TGCTGGGAAA AGACAGGAGA AGTACTGACT CAACTGCACT GACCATGTTG
1651 TCATAATTAG AATAAAGAAG AAGTGGTCGG AAATGCACAT TCCTGGATAG
1701 GAATCACAGC TCACCCCAGG ATCTCACAGG TAGTCTCCTG AGTAGTTGAC
1751 GGCTAGCGGG GAGCTAGTTC CGCCGCATAG TTATAGTGTT GATGTGTGAA
1801 CGCTGACCTG TCCTGTGTGC TAAGAGCTAT GCAGCTTAGC TGAGGCGCCT
1851 AGATTACTAG ATGTGCTGTA TCACGGGGAA TGAGGTGGGG GTGCTTATTT
1901 TTTAATGAAC TAATCAGAGC CTCTTGAGAA ATTGTTACTC ATTGAACTGG
1951 AGCATCAAGA CATCTCATGG AAGTGGATAC GGAGTGATTT GGTGTCCATG
2001 CTTTTCACTC TGAGGACATT TAATCGGAGA ACCTCCTGGG GAATTTTGTG
2051 GGAGACACTT GGGAACAAAA CAGACACCCT GGGAATGCAG TTGCAAGCAC
2101 AGATGCTGCC ACCAGTGTCT CTGACCACCC TGGTGTGACT GCTGACTGCC
2151 AGCGTGGTAC CTCCCATGCT GCAGGCCTCC ATCTAAATGA GACAACAAAG
2201 CACAATGTTC ACTGTTTACA ACCAAGACAA CTGCGTGGGT CCAAACACTC
2251 CTCTTCCTCC AGGTCATTTG TTTTGCATTT TTAATGTCTT TATTTTTTGT
2301 AATGAAAAAG CACACTAAGC TGCCCCTGGA ATCGGGTGCA GCTGAATAGG
2351 CACCCAAAAG TCCGTGACTA AATTCCGTTT GTCTTTTTGA TAGCAAATTA
2401 TGTTAAGAGA CAGTGATGGC TAGGGCTCAA CAATTTTGTA TTCCCATGTT
2451 TGTGTGAGAC AGAGTTTGTT TTCCCTTGAA CTTGGTTAGA ATTGTGCTAC
2501 TGTGAACGCT GATCCTGCAT ATGGAAGTCC CACTTTGGTG ACATTTCCTG
2551 GCCATTCTTG TTTCCATTGT GTGGATGGTG GGTTGTGCCC ACTTCCTGGA
2601 GTGAGACAGC TCCTGGTGTG TAGAATTCCC GGAGCGTCCG TGGTTCAGAG
2651 TAAACTTGAA GCAGATCTGT GCATGCTTTT CCTCTGCAGC AATTGGCTCG
2701 TTTCTCTTTT TTGTTCTCTT TTGATAGGAT CCTGTTTCCT ATGTGTGCAA
```

2751 AATAAAAATA AATTTGGGCA AAAAAAAAAA AAAAAAAAA

BLAST Results

Entry HS671255 from database EMBL:
human STS SHGC-11828.
Length = 400
Minus Strand HSPs:
Score = 1822 (273.4 bits), Expect = 4.8e-76, P = 4.8e-76
Identities = 382/397 (96%), Positives = 382/397 (96%),

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 10 bp to 525 bp; peptide length: 172
Category: putative protein
Classification: unset

```
  1 MRRQPAKVAA LLLGLLLECT EAKKHCWYFE GLYPTYYICR SYEDCCGSRC
 51 CVRALSIQRL WYFWFLLMMG VLFCCGAGFF IRRRMYPPPL IEEPAFNVSY
101 TRQPPNPGPG AQQPGPPYYT DPGGPGMNPV GNSTAMAFQV PPNSPQGSVA
151 CPPPPAYCNT PPPPYEQVVK AK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_22k8, frame 1

PIR:S14970 extensin class I (clone w17-1) - tomato, N = 1, Score = 118,
P = 2.3e-07

>PIR:S14970 extensin class I (clone w17-1) - tomato
Length = 132

HSPs:

Score = 118 (17.7 bits), Expect = 2.3e-07, P = 2.3e-07
Identities = 30/82 (36%), Positives = 35/82 (42%)

```
Query:    87 PPPLIEEPAFNVSYTRQPPNPGPGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQ 146
             PPP     P       Y + PP P P      P P YY  P  P  +P         +  P  SP
Sbjct:    32 PPPSPSPPP--PYYYKSPPPPSPSP--PPPYYYKSPPPPDPSPPPPYYYKSPPPPSPSPP 87

Query:   147 GSVACPPPPAYCNTPPPP--YEQV 168
                    PPPP Y + PPPP  YE +
Sbjct:    88 PPSPSPPPPTYSSPPPPPPFYENI 111
```

Score = 104 (15.6 bits), Expect = 6.9e-06, P = 6.9e-06
Identities = 28/78 (35%), Positives = 34/78 (43%)

```
Query:    87 PPPLIEEPAFNVSYTRQPPNPGPGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQ 146
             PP      P +   Y + PP P P      P P YY  P  P  +P           ++ PP  P
Sbjct:     1 PPSPSPPPPY---YYKSPPPPSPSP--PPPYYYKSPPPPSPSP---PPPYYYKSPP-PPS 51

Query:   147 GSVACPPPPAYCNTPPPP 164
             S    PPPP Y  +PPPP
Sbjct:    52 PS---PPPPYYYKSPPPP 66
```

Score = 102 (15.3 bits), Expect = 1.1e-05, P = 1.1e-05
Identities = 30/78 (38%), Positives = 33/78 (42%)

```
Query:    87 PPPLIEEPAFNVSYTRQPPNPGPGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQ 146
             PPP     P       Y + PP P P      P P YY  P  P  +P  S +     PP  P
Sbjct:    48 PPPSPSPPP--PYYYKSPPPPDPSP--PPPYYYKSPPPPSPSPPPPSPS-----PP-PPT 97
```

```
Query:   147 GSVACPPPPAYCNTPPPP 164
             S   PPPP Y N P PP
Sbjct:    98 YSSPPPPPPFYENIPLPP 115

 Score = 95 (14.3 bits), Expect = 2.4e-04, P = 2.4e-04
 Identities = 24/61 (39%), Positives = 29/61 (47%)

Query:   104 PPNPGPGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPNSPQGSVACPPPPAYCNTPPP 163
             PP+P P    P P YY  P  P  +P         ++ PP  P S    PPPP Y  +PPP
Sbjct:     1 PPSPSP----PPPYYYKSPPPPSPSP---PPPYYYKSPP-PPSPS---PPPPYYYKSPPP 49

Query:   164 P 164
             P
Sbjct:    50 P 50

 Score = 68 (10.2 bits), Expect = 4.2e+00, P = 9.8e-01
 Identities = 24/69 (34%), Positives = 29/69 (42%)

Query:    87 PPPLIEEPAFNVSYTRQPP---NPGPGAQQPGPPYYTDPGGPGMNPVGNSTAMAFQVPPN 143
             PPP     P     Y  PP   +P P +  P PP Y+ P  P   P     +  +    PP
Sbjct:    63 PPPPDPSPPPPYYYKSPPPPSPSPPPPSPSPPPPTYSSPPPPP--PFYENIPL----PPV 116

Query:   144 SPQGSVACPPPP 155
                  S A PPPP
Sbjct:   117 IGV-SYASPPPP 127
```

Peptide information for frame 3

```
ORF from 0 bp to 368 bp; peptide length: 123
Category: questionable ORF
Classification: unset


      1 GSHEAPACEG GGAAARAALG VHRSQKALLV FRRTLSNLLY MPLLRGLLWL
     51 QVLCAGPLHT EAVVLLVPSD DGRAFLLRSR LLHPEAHVPP AADRGASLQC
    101 VLHQAAPKSR PRSPAAGAAL LH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_22k8, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_22k8, frame 1

Report for DKFZphfbr2_22k8.1

```
[LENGTH]        172
[MW]            19194.47
[pI]            8.77
[KW]            SIGNAL_PEPTIDE 23
[KW]            TRANSMEMBRANE  1
[KW]            LOW_COMPLEXITY     27.33 %

SEQ     MRRQPAKVAALLLGLLLECTEAKKHCWYFEGLYPTYYICRSYEDCCGSRCCVRALSIQRL
SEG     ..........xxxxxxx...........................................
PRD     ccchhhhhhhhhhhhhhhhhhhhhhcccccccccceeeecccccccccccchhhhhhhhh
MEM     ............................................................

SEQ     WYFWFLLMMGVLFCCGAGFFIRRRMYPPPLIEEPAFNVSYTRQPPNPGPGAQQPGPPYYT
SEG     ..........................................xxxxxxxxxxxxxxxxxx
PRD     hhhhhhhhhhhhhcccccceeeeecccccccccccccccceeeecccccccccccccccccc
MEM     ....MMMMMMMMMMMMMMMMM.......................................

SEQ     DPGGPGMNPVGNSTAMAFQVPPNSPQGSVACPPPPAYCNTPPPPYEQVVKAK
SEG     xxxxxx.......................xxxxxxxxxxxxxxxx.......
PRD     cccccccccccccceeecccccccccccccccccccccccccccccccccccc
MEM     ....................................................
```

(No Prosite data available for DKFZphfbr2_22k8.1)

(No Pfam data available for DKFZphfbr2_22k8.1)

Pedant information for DKFZphfbr2_22k8, frame 3

Report for DKFZphfbr2_22k8.3

```
[LENGTH]        122
[MW]            12854.08
[pI]            10.27
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY     25.41 %

SEQ     GSHEAPACEGGGAAARAALGVHRSQKALLVFRRTLSNLLYMPLLRGLLWLQVLCAGPLHT
SEG     ....xxxxxxxxxxxxxxxx........................................
PRD     ccccccccccchhhhhhhccccchhhhhhhhhhhhhhhhcccccchhhhhhhhhcccccc

SEQ     EAVVLLVPSDDGRAFLLRSRLLHPEAHVPPAADRGASLQCVLHQAAPKSRPRSPAAGAAL
SEG     ............................................xxxxxxxxxxxxxxx.
PRD     cceeeeecccccchhhhhhhhcccccccccccccchhhhhhhhhcccccccccchhhhhhc

SEQ     LH
SEG     ..
PRD     cc
```

(No Prosite data available for DKFZphfbr2_22k8.3)

(No Pfam data available for DKFZphfbr2_22k8.3)

DKFZphfbr2_23b10
----

group: nucleic acid managment

DKFZphfbr2_2b10 encodes a novel 580 amino acid protein with strong similarity to rat RNA helicase HEL117.

HEL117 is a DEAD/H box helicase, which co-localises with a splicing factor and thus seems to be involved in splicing.

The new protein can find application in modulation of splicing.

strong similarity to rat RNA helicase HEL117

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 2905 bp
Poly A stretch at pos. 2885, no polyadenylation signal found

```
   1 GGGGGCTCCG CTCCGCACCA CCAACCCCGG GCCGCAGTCC TGACGAGCGG
  51 GTCAGGGCTT GTCGGGCGGA AGCCTGGCCT GGAGCCTGGA AGGGGGAGAC
 101 GGCCCGAGCG GGAGCGGGAG CGGACGCGGC CTCAGTCCTG CGCGGAATAT
 151 TGAAGGATGT TTGTTCCAAG ATCTCTAAAA ATCAAGAGGA ATGCTAATGA
 201 TGATGGCAAA AGTTGTGTGG CTAAGATAAT TAAACCAGAC CCAGAAGACC
 251 TTCAGTTGGA CAAAAGCAGA GATGTTCCCG TTGATGCTGT AGCTACAGAA
 301 GCAGCCACAA TAGACAGGCA CATCAGCGAA TCATGCCCTT TCCCCAGCCC
 351 AGGTGGCCAG TTGGCAGAGG TTCATTCAGT AAGTCCCGAG CAGGGTGCGA
 401 AGGACAGCCA TCCTTCTGAA GAGCCCGTTA AGTCATTTTC CAAAACACAG
 451 CGCTGGGCAG AACCAGGGGA ACCCATCTGT GTTGTCTGTG GTCGTTATGG
 501 AGAGTATATC TGTGATAAGA CAGATGAAGA TGTGTGTAGT TTGGAGTGTA
 551 AAGCGAAACA TCTTCTACAA GTTAAGGAAA AGGAAGAGAA ATCAAAACTC
 601 AGCAATCCAC AGAAGGCTGA TTCTGAGCCA GAGTCTCCAC TGAATGCTTC
 651 CTATGTCTAC AAAGAGCACC CCTTTATTTT GAACCTTCAG GAAGACCAGA
 701 TTGAAAATCT TAAACAGCAG CTGGGAATTT TAGTTCAAGG GCAAGAAGTC
 751 ACCAGGCCCA TTATTGACTT TGAACATTGT AGTCTCCCTG AGGTCTTAAA
 801 TCACAACTTG AAGAAATCAG GCTATGAGGT GCCAACTCCC ATTCAAATGC
 851 AGATGATTCC TGTGGGACTT CTGGGAAGAG ACATTCTGGC CAGTGCAGAT
 901 ACTGGCTCAG GAAAAACAGC TGCTTTTCTT CTTCCTGTTA TCATGCGAGC
 951 TTTATTCGAG AGCAAAACTC CATCTGCGCT CATTCTTACA CCAACCAGAG
1001 AGTTAGCCAT TCAGATAGAG AGACAAGCTA AAGAATTGAT GAGTGGCCTG
1051 CCACGCATGA AAACTGTGCT TCTTGTAGGG GGCTTACCCT TACCCCCACA
1101 GCTTTATCGT CTGCAACAAC ATGTTAAGGT TATCATAGCA ACCCCTGGGC
1151 GACTTCTGGA TATAATAAAG CAGAGCTCTG TAGAACTCTG TGGTGTAAAG
1201 ATTGTGGTAG TAGATGAAGC TGATACCATG TTAAAGATGG GTTTTCAACA
1251 ACAAGTGCTT GACATTTTGG AAAACATTCC TAATGATTGT CAGACCATTT
1301 TGGTTTCAGC CACAATTCCA ACTAGCATAG AACAGCTAGC AAGCCAGCTT
1351 CTGCATAATC CTGTGAGAAT TATCACTGGA GAAAAGAACC TACCTTGTGC
1401 CAATGTACGT CAGATTATTT TGTGGGTAGA AGACCCAGCC AAAAAGAAAA
1451 AATTATTTGA AATTTTAAAT GATAAGAAAC TCTTTAAGCC TCCAGTGTTA
1501 GTATTTGTGG ACTGCAAACT AGGAGCAGAT CTTTTGAGTG AAGCCGTTCA
1551 GAAAATCACA GGGCTGAAAA GCATATCTAT ACATTCGGAG AAGTCGCAAA
1601 TAGAAAGGAA AAACATATTG AAGGGATTAC TTGAAGGAGA CTATGAAGTT
1651 GTAGTGAGCA CAGGAGTCTT GGGACGAGGC CTAGACTTGA TCAGTGTCAG
1701 GCTGGTTGTC AATTTTGATA TGCCTTCAAG TATGGATGAG TATGTCCATC
1751 AGGAAAATAC CTACAAGTCT ACTTGGAGGA ATCCCCAGCA TTTTCAACAG
1801 GATGTCAGAA TGACCTTGGG CTATGTTGGC AAAGCACAAT GGGAAGAAGA
1851 CAACCAATTG AAGGTCAAAC TAGGCCTTAA AAAAAATTGT TCTTCCTAAA
1901 TGAAACTTTA TGTAAGACCC AAGCTTCCTT TATGTAAAAA TAGGATACTC
1951 ACTAGGCTTT GGGGCTGACA ATGGTTTTTA AATCTTGCTA ATCTTCCCTG
2001 GAATGAAACC AGCATGACTC AAAGAGAAAA AGAGAGTCTA TAATATTTTC
2051 TAATCCCTGA GTTCTTTTCT TTATATATTA AAAAGGATTA TTAGGCTGGG
2101 TGTGGTGGCT CACGCCTGTA ATCCCAGCAC TTTGGGAGGC CGAGGGGAGT
2151 GGATCACCTG AGTTCGAGAC CAGCCTAACC AACATGGAGA AACCCTGTCT
2201 CTACTAAAAA TACAAAATTA GCCAGGCGTG GTGGCGCATG CCTGTAATCC
2251 CAGCTACTCA GGAGGCTACA GCAGGAGAAT TGCTTGAACT CGGGAGGCAG
2301 AGCCAAGATC GCACCACTGC ACTCCAGCCT GGGCAACAAG AGTGAAACTC
2351 TGTCTCAAAA TAATATTAAT GATAATAATA ATAATAATAA TAGGGATTAC
2401 TTGCATAATT GTTCTTTTAA AATTATTGGC AGTATTGCTG AATGTATTTA
2451 GATTTTTTCA CCAAGTGACA ACAACTGAAT TCATAAAGAT TCATCAACAA
2501 GACCTGATAA AAAAAAATGT AAGCATATTA TAGTGGATAC TTCCAAGACT
2551 CTTGGTCTAA CATGTATTAG AAAGCAGAAG GAGCCCAGGC ACAGGGGCTC
2601 CCGCCGGTAA TCCCAAAGCT TTGGGAAGCC AAGGCAGGTG GATCGCTTGA
2651 GCTCAGGAGT TAGAGACCAG CCTGGGCAAC ATGGTGAAAT CCCGTCACCA
```

```
2701 CAAAAAAATG CAAAAATTAA CTGGGCGTGG TGGCATGCAC CTGTAGTCCC
2751 AGCTACTCTG GAGGCTGAGG TGAGGGGAAT CACCTGAGCC GGGGGAATCA
2801 CCTGAGCCCA GGGAAGTTGA GGCTGCTGTG AGCCATGGTC ATGACACTGC
2851 CCTCCAGCCT GGACAACAGA TTGAGACCCT GTCTCAAAAA AAAAAAAAAA
2901 AAAAA
```

BLAST Results

No BLAST result

Medline entries

Medline:
A putative mammalian RNA helicase with an arginine-serine-rich domain

Peptide information for frame 1

ORF from 157 bp to 1896 bp; peptide length: 580
Category: strong similarity to known protein
Prosite motifs: ATP_GTP_A (247-255)
LEUCINE_ZIPPER (298-320)

```
  1 MFVPRSLKIK RNANDDGKSC VAKIIKPDPE DLQLDKSRDV PVDAVATEAA
 51 TIDRHISESC PFPSPGGQLA EVHSVSPEQG AKDSHPSEEP VKSFSKTQRW
101 AEPGEPICVV CGRYGEYICD KTDEDVCSLE CKAKHLLQVK EKEEKSKLSN
151 PQKADSEPES PLNASYVYKE HPFILNLQED QIENLKQQLG ILVQGQEVTR
201 PIIDFEHCSL PEVLNHNLKK SGYEVPTPIQ MQMIPVGLLG RDILASADTG
251 SGKTAAFLLP VIMRALFESK TPSALILTPT RELAIQIERQ AKELMSGLPR
301 MKTVLLVGGL PLPPQLYRLQ QHVKVIIATP GRLLDIIKQS SVELCGVKIV
351 VVDEADTMLK MGFQQQVLDI LENIPNDCQT ILVSATIPTS IEQLASQLLH
401 NPVRIITGEK NLPCANVRQI ILWVEDPAKK KKLFEILNDK KLFKPPVLVF
451 VDCKLGADLL SEAVQKITGL KSISIHSEKS QIERKNILKG LLEGDYEVVV
501 STGVLGRGLD LISVRLVVNF DMPSSMDEYV HQENTYKSTW RNPQHFQQDV
551 RMTLGYVGKA QWEEDNQLKV KLGLKKNCSS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23b10, frame 1

PIR:A57514 RNA helicase HEL117 - rat, N = 2, Score = 615, P = 1.6e-60

TREMBL:AB018344_1 gene: "KIAA0801"; product: "KIAA0801 protein"; Homo sapiens mRNA for KIAA0801 protein, complete cds., N = 1, Score = 615, P = 2.8e-59

TREMBL:CEF01F1_1 gene: "F01F1.7"; Caenorhabditis elegans cosmid F01F1., N = 2, Score = 365, P = 1.9e-58

TREMBL:AF083255_1 product: "RNA helicase-related protein"; Homo sapiens RNA helicase-related protein mRNA, complete cds., N = 2, Score = 556, P = 1.5e-57

PIR:S14048 RNA helicase dbp2 - fission yeast (Schizosaccharomyces pombe), N = 1, Score = 591, P = 1.6e-57

>PIR:A57514 RNA helicase HEL117 - rat
Length = 1,032

HSPs:

Score = 615 (92.3 bits), Expect = 1.6e-60, Sum P(2) = 1.6e-60
Identities = 140/394 (35%), Positives = 236/394 (59%)

```
Query:   144 EKSKLSNPQKADSEPESPLNASYVYKEHPFILNLQEDQIENLKQQL-GILVQGQEVTRPI 202
             ++ KL  P          P  ++ Y E P +  + ++++   + ++ GI V+G+   +PI
Sbjct:   313 KQRKLLEPVDHGKIEYEPFRKNF-YVEVPELAKMSQEEVNVFRLEMEGITVKGKGCPKPI 371

Query:   203 IDFEHCSLPEVLNHNLKKSGYEVPTPIQMQMIPVGLLGRDILASADTGSGKTAAFLLPV- 261
```

```
                 +  C +    + ++LKK GYE PTPIQ Q IP   + GRD++  A TGSGKT AFLLP+
Sbjct:   372 KSWVQCGISMKILNSLKKHGYEKPTPIQTQAIPAIMSGRDLIGIAKTGSGKTIAFLLPMF 431

Query:   262 --IM--RALFESKTPSALILTPTRELAIQIERQAKELMSGLPRMKTVLLVGGLPLPPQLY 317
               IM  R+L E + P A+I+TPTRELA+QI ++ K+    L  ++ V + GG  +  Q+
Sbjct:   432 RHIMDQRSLEEGEGPIAVIMTPTRELALQITKECKKFSKTLG-LRVVCVYGGTGISEQIA 490

Query:   318 RLQQHVKVIIATPGRLLDIIKQSS---VELCGVKIVVVDEADTMLKMGFQQQVLDILENI 374
             L++  ++I+ TPGR++D++  +S       L  V  VV+DEAD M  MGF+ QV+ I++N+
Sbjct:   491 ELKRGAEIIVCTPGRMIDMLAANSGRVTNLRRVTYVVLDEADRMFDMGFEPQVMRIVDNV 550

Query:   375 PNDCQTILVSATIPTSIEQLASQLLHNPVRIITGEKNLPCANVRQIILWVEDPAKKKKLF 434
               D QT++ SAT P ++E LA ++L  P+ +  G +++ C++V Q ++ +E+  K KL
Sbjct:   551 RPDRQTVMFSATFPRAMEALARRILSKPIEVQVGGRSVVCSDVEQQVIVIEEEKKFLKLL 610

Query:   435 EILNDKKLFKPPVLVFVDCKLGADLLSEAVQKITGLKSISIHSEKSQIERKNILKGLLEG 494
             E+L    +     V++FVD +  AD L + + + +     +S+H     Q +R +I+     G
Sbjct:   611 ELLGHYQE-SGSVIIFVDKQEHADGLLKDLMRAS-YPCMSLHGGIDQYDRDSIINDFKNG 668

Query:   495 DYEVVVSTGVLGRGLDLISVRLVVNFDMPSSMDEYVHQ 532
             +++V+T V  RGLD+  + LVVN+  P+  ++YVH+
Sbjct:   669 TCKLLVATSVAARGLDVKHLILVVNYSCPNHYEDYVHR 706

 Score = 37 (5.6 bits), Expect = 1.6e-60, Sum P(2) = 1.6e-60
 Identities = 13/36 (36%), Positives = 17/36 (47%)

Query:   132 KAKHLLQVKEKEE---KSKLSNPQKADSEPESPLNA 164
             KA++  + KEK E     SK     K D E E  +A
Sbjct:   113 KAENRSRSKEKAEGGDSSKEKKKDKDDKEDEKEKDA 148
```

Pedant information for DKFZphfbr2_23b10, frame 1

Report for DKFZphfbr2_23b10.1

```
[LENGTH]        580
[MW]            64572.24
[pI]            6.13
[HOMOL]         TREMBL:CEF01F1_1 gene: "F01F1.7"; Caenorhabditis elegans cosmid F01F1. 8e-61

[FUNCAT]        30.10 nuclear organization     [S. cerevisiae, YNL112w] 2e-53
[FUNCAT]        04.01.04 rrna processing       [S. cerevisiae, YNL112w] 2e-53
[FUNCAT]        04.05.03 mrna processing (splicing)   [S. cerevisiae, YPL119c] 5e-53
[FUNCAT]        30.03 organization of cytoplasm      [S. cerevisiae, YOR204w] 2e-49
[FUNCAT]        05.04 translation (initiation, elongation and termination)  [S. cerevisiae,
YOR204w] 2e-49
[FUNCAT]        j mrna translation and ribosome biogenesis    [H. influenzae, HI0231 RNA] 2e-46
[FUNCAT]        06.10 assembly of protein complexes   [S. cerevisiae, YLL008w] 3e-43
[FUNCAT]        04.99 other transcription activities  [S. cerevisiae, YDL160c] 4e-39
[FUNCAT]        l genome replication, transcription, recombination and repair      [H.
influenzae, HI0892] 3e-35
[FUNCAT]        04.05.01.07 chromatin modification    [S. cerevisiae, YMR290c] 6e-34
[FUNCAT]        98 classification not yet clear-cut   [S. cerevisiae, YOR046c] 3e-32
[FUNCAT]        09.01 biogenesis of cell wall         [S. cerevisiae, YJL033w] 8e-30
[FUNCAT]        30.16 mitochondrial organization      [S. cerevisiae, YDR194c] 5e-23
[FUNCAT]        99 unclassified proteins    [S. cerevisiae, YGL064c] 1e-16
[FUNCAT]        r general function prediction        [M. jannaschii, MJ1401] 5e-11
[FUNCAT]        11.10 cell death       [S. cerevisiae, YMR190c] 1e-06
[FUNCAT]        03.19 recombination and dna repair    [S. cerevisiae, YMR190c] 1e-06
[BLOCKS]        BL00115B Eukaryotic RNA polymerase II heptapeptide repeat proteins
[BLOCKS]        BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]        BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]        BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]        BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW]         nucleus 6e-53
[PIRKW]         RNA binding 9e-52
[PIRKW]         DEAD box 2e-43
[PIRKW]         transmembrane protein 1e-21
[PIRKW]         DNA binding 5e-48
[PIRKW]         ATP 4e-57
[PIRKW]         purine nucleotide binding 2e-43
[PIRKW]         P-loop 4e-57
[PIRKW]         hydrolase 6e-42
[PIRKW]         protein biosynthesis 2e-43
[PIRKW]         ATP binding 2e-50
[SUPFAM]        WW repeat homology 1e-49
[SUPFAM]        translation initiation factor eIF-4A 2e-43
[SUPFAM]        DEAD/H box helicase homology 4e-57
[SUPFAM]        recQ helicase homology 8e-06
```

```
[SUPFAM]        unassigned DEAD/H box helicases 4e-57
[SUPFAM]        ATP-dependent RNA helicase DBP1 2e-53
[SUPFAM]        ATP-dependent RNA helicase DHH1 6e-40
[SUPFAM]        tobacco ATP-dependent RNA helicase DB10 1e-49
[SUPFAM]        Bloom's syndrome helicase 8e-06
[PROSITE]       ATP_GTP_A       1
[PROSITE]       LEUCINE_ZIPPER 1
[PROSITE]       MYRISTYL        6
[PROSITE]       CK2_PHOSPHO_SITE        8
[PROSITE]       TYR_PHOSPHO_SITE        1
[PROSITE]       PKC_PHOSPHO_SITE        7
[PROSITE]       ASN_GLYCOSYLATION       1
[PFAM]          Helicases conserved C-terminal domain
[PFAM]          DEAD and DEAH box helicases
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY      3.10 %

SEQ   MFVPRSLKIKRNANDDGKSCVAKIIKPDPEDLQLDKSRDVPVDAVATEAATIDRHISESC
SEG   ............................................................
PRD   cccccceeeccccccccceeeeeeeeccccceeecccccccccchhhhhhhhhhhhcccc

SEQ   PFPSPGGQLAEVHSVSPEQGAKDSHPSEEPVKSFSKTQRWAEPGEPICVVCGRYGEYICD
SEG   ............................................................
PRD   cccccccceeeccccccccccccccccccccccccccccccccccceeeecccccceeec

SEQ   KTDEDVCSLECKAKHLLQVKEKEEKSKLSNPQKADSEPESPLNASYVYKEHPFILNLQED
SEG   ............................................................
PRD   cccccccchhhhhhhhhhhhhhccccccccccccccccccccccceeeccccccccchhh

SEQ   QIENLKQQLGILVQGQEVTRPIIDFEHCSLPEVLNHNLKKSGYEVPTPIQMQMIPVGLLG
SEG   ............................................................
PRD   hhhhhhhhheeeecccccccccccccccccchhhhhhhhhhcccccccccccccceeeecc

SEQ   RDILASADTGSGKTAAFLLPVIMRALFESKTPSALILTPTRELAIQIERQAKELMSGLPR
SEG   ............................................................
PRD   cceeeeecccccccceeeehhhhhhhhcccccceeeeecchhhhhhhhhhhhhhhhccccc

SEQ   MKTVLLVGGLPLPPQLYRLQQHVKVIIATPGRLLDIIKQSSVELCGVKIVVVDEADTMLK
SEG   ...xxxxxxxxxxxxxxxxx........................................
PRD   eeeeeeecccccchhhhhhhheeeeeeccccchhhhhhheeeeeeeeeeeehhhhhhhh

SEQ   MGFQQQVLDILENIPNDCQTILVSATIPTSIEQLASQLLHNPVRIITGEKNLPCANVRQI
SEG   ............................................................
PRD   ccccchhhhhhhhhcccccceeeeeccccchhhhhhhhhhhceeeeeeeccccccccccce

SEQ   ILWVEDPAKKKKLFEILNDKKLFKPPVLVFVDCKLGADLLSEAVQKITGLKSISIHSEKS
SEG   ............................................................
PRD   eeeccccchhhhhhhhhhhhcccccceeeeeeeccccchhhhhhhhhhhccceeeccccch

SEQ   QIERKNILKGLLEGDYEVVVSTGVLGRGLDLISVRLVVNFDMPSSMDEYVHQENTYKSTW
SEG   ............................................................
PRD   hhhhhhhhhhhccccceeeeehhhhhhccccceeeeeeeeccccccccceeeeccccccc

SEQ   RNPQHFQQDVRMTLGYVGKAQWEEDNQLKVKLGLKKNCSS
SEG   ........................................
PRD   ccccccchhhhhhhccccchhhhhhhhhhhhhhhcccccc
```

Prosite for DKFZphfbr2_23b10.1

PS00001	163->167	ASN_GLYCOSYLATION	PDOC00001
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDOC00005
PS00005	251->254	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	513->516	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	539->542	PKC_PHOSPHO_SITE	PDOC00005
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00006	389->393	CK2_PHOSPHO_SITE	PDOC00006
PS00006	480->484	CK2_PHOSPHO_SITE	PDOC00006
PS00006	524->528	CK2_PHOSPHO_SITE	PDOC00006
PS00007	489->497	TYR_PHOSPHO_SITE	PDOC00007
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	80->86	MYRISTYL	PDOC00008

PS00008	195->201	MYRISTYL	PDOC00008
PS00008	250->256	MYRISTYL	PDOC00008
PS00008	490->496	MYRISTYL	PDOC00008
PS00008	573->579	MYRISTYL	PDOC00008
PS00017	247->255	ATP_GTP_A	PDOC00017
PS00029	298->320	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphfbr2_23b10.1

```
HMM_NAME    DEAD and DEAH box helicases

HMM             *gLpPWILRnIyeMGFEkPTPIQQqAIPiILeGRDVMACAQTGSGKTAAF
                  +LP+ +   N+++ G+E PTPIQ+Q IP+ L GRD++A A TGSGKTAAF
Query       209   SLPEVLNHNLKKSGYEVPTPIQMQMIPVGLLGRDILASADTGSGKTAAF    257

HMM             lIPMLQHIDwdPWpqpPQdPrALILAPTRELAMQIQEEcRkFgkHMngIR
                L+P++ + + +    ++P     ALIL+PTRELA+QI+++++++ + ++ ++
Query       258 LLPVIMRALFES--KTPS---ALILTPTRELAIQIERQAKELMSGLPRMK    302

HMM             ImcIYGGtnMRdQMRmLeRGpPHIVIATPGRLIDHIERgtldLDrIeMLV
                 ++++GG+++ +Q+ +L++ +  ++IATPGRL+D+I++ ++ L  ++++V
Query       303 TVLLVGGLPLPPQLYRLQQHV-KVIIATPGRLLDIIKQSSVELCGVKIVV    351

HMM             MDEADRMLDMGFIDQIRrIMrqIPMpwNRQTMMFSATMPdeIqELARrFM
                 DEAD ML MGF++Q+ +I+  IP    + QT++ SAT+P +I++LA  ++
Query       352 VDEADTMLKMGFQQQVLDILENIP--NDCQTILVSATIPTSIEQLASQLL    399

HMM             RNPIRInIdMdElTtnEnIkQwYiyVerEMWKfdcLcrLIe*
                +NP+RI+ ++++L    N++Q++ +VE +  K  +L+++++
Query       400 HNPVRIITGEKNLPCA-NVRQIILWVE-DPAKKKKLFEILN    438


HMM_NAME    Helicases conserved C-terminal domain

HMM             *EileeWLknl.GIrvmYIHGdMpQeERdeIMddFNnGEynVLIcTDVgg
                 ++L+E  ++  G++ ++IH+ ++Q ER +I++   +G+Y V ++T V+G
Query       458  DLLSEAVQKITGLKSISIHSEKSQIERKNILKGLLEGDYEVVVSTGVLG    506

HMM             RGIDIPdVNHVINYDMPWNPEqYIQRIGRTgRIG*
                RG+D+++V++V+N+DMP +++ Y++ +  T  +
Query       507 RGLDLISVRLVVNFDMPSSMDEYVH-QENTYKST    539
```

DKFZphfbr2_23b21
----

group: signal transduction

DKFZphfbr2_23b21.1 encodes a novel 193 amino acid protein which is nearly identical to bovine neurocalcin.

Neurocalcin is a Ca(2+)-binding protein with three putative Ca(2+)-binding domains (EF-hands). In cattle, 6 isoforms are differentially expressed in the central nervous system, retina and adrenal gland. Homology with recoverin indicates involvement in Ca2+ dependent activation of guanylate cyclase.

The new protein can find application in modulating/blocking the guanylate cyclase-pathway.

nearly identical to bovine neurocalcin

complete cds complete cDNA
EST hits

Sequenced by AGOWA

Locus: /map="574.6 cR from top of Chr8 linkage group"

Insert length: 3300 bp
Poly A stretch at pos. 3279, polyadenylation signal at pos. 3249

```
   1 GGGGAGAATC TGGTGGATGC TGGACCTTGC TGCTGCTGCT ACTGCTGTTT
  51 CCAGGGGCTG CAGAGCATGG ACTGTTAAAT CTTGCACTTC TTCTGAGTGA
 101 GCTGAATTCT TGCCGCCAGG ATGGGGAAAC AGAACAGCAA GCTGCGCCCG
 151 GAGGTCATGC AGGACTTGCT GGAAAGCACA GACTTTACAG AGCATGAGAT
 201 CCAGGAATGG TATAAAGGCT TCTTGAGAGA CTGCCCCAGT GGACATTTGT
 251 CAATGGAAGA GTTTAAGAAA ATATATGGGA ACTTTTTCCC TTATGGGGAT
 301 GCTTCCAAAT TTGCAGAGCA TGTCTTCCGC ACCTTCGATG CAAATGGAGA
 351 TGGGACAATA GACTTTAGAG AATTCATCAT CGCCTTGAGT GTAACTTCGA
 401 GGGGGAAGCT GGAGCAGAAG CTGAAATGGG CCTTCAGCAT GTACGACCTG
 451 GACGGAAATG GCTATATCAG CAAGGCAGAG ATGCTAGTGA TCGTGCAGGC
 501 AATCTATAAG ATGGTTTCCT CTGTAATGAA AATGCCTGAA GATGAGTCAA
 551 CCCCAGAGAA AAGAACAGAA AAGATCTTCC GCCAGATGGA CACCAATAGA
 601 GACGGAAAAC TCTCCCTGGA AGAGTTCATC CGAGGAGCCA AAAGCGACCC
 651 GTCCATTGTG CGCCTCCTGC AGTGCGACCC GAGCAGTGCC GGCCAGTTCT
 701 GAGCCCTGCG CCCACCAATC GAATTGTAGA GCTGCTTGTG TTCCCTTTTG
 751 ATTCTTCTTT TTAACAATTT TTTTTTTTTT TTGCCAAACA ATATCAATGG
 801 TGATGCCGTC CCCTGTGCGG TCTGATGCGC CTTCCTCCGT GACGCCTTCA
 851 GCCTCTTTTG TCGTGGATGC TTCGTGGGAA TGCCCAGAGC CCCAGTGTGC
 901 TTGTGGAGAG CATGGACAGA CTTCGTGGTG TTCATTGTTT GATGATTTTT
 951 AATCGTTACT ATTATTTCTT TTTATTCTAA TGTCTCTGTT CTAAAACGTA
1001 AGACTCGGGG GTTGGGGCAA AAGAAGGGAA ACCCATCCAG TCCTGTGATT
1051 CTATTGCAAG CTTCAAGGGG CTTTTGTTTG AAAGACAAAA CTCCCCACCT
1101 GGGTCTGTTG TCACACGTGC CGTAGGGGTG ATGGATGGCA CCGGATGCTG
1151 GATTCCCCAA GAACAAGTTA CCCTCTGGGG TGAGGCTATT CCAGCGAGCT
1201 GGGACATTTC CCCATGGGGG CCCACTCCCC TCTCTTCCCC AGCAGGCTGT
1251 AGTTTCTAAG CTGTGAACAT TTCAAGATAA ATTAACAGAG GAGAGGAAAA
1301 AGATGGCTCA GCTATTTTTT CACAGGTTTA CACTAGTTGA GCTAATATGC
1351 GTGTCTTTGG AAATTAAACA CAAATGGTAA CATATTCCAA AACCAGACCC
1401 ATCTTGTTGC CTATTGTGAT AAAATAAAAA GACGGCTGTA TATAACATAT
1451 TGGGTAATGC AGACCAAATT AAGTGTTTTG CCTTGTTTAA ATGAAATGCA
1501 TGTTTAGTGA GCACTAATAC AATCTTATTC CAGAAGACTG TTTTTAGTAG
1551 CTTATTGTGA AGTAAGACAA CTATAATGAA TGTCTGTCTT GTTTGGAAGT
1601 CATATCTGTC TTTGCACAAA TGTACCAATC GACAAGTATA TTTTATATAT
1651 TCCATAAAAA TACAAAGTAA CCCTGACTAG GGCCCAACTT TAATTTTGAA
1701 TGCATTTCCA GAGTGGCCAT GCCTAGAGGG CAGATGCAGA GCAGGTGGTA
1751 GTGGGACAGG ACAATTGGAG CACAGGAATG TTAACATGTA TGACAGGGGA
1801 CCAGTAGGGT GGTTTCCCTC TCAGGCCCAG CAGCCCATTG ACAGCATTAG
1851 ACTGGCGGCA TGGTGCTTTT CTGAGCAGAT CAATACTCTG CAGACTCGAA
1901 AAAACATCAC ATACATTCTT GGAACTTCCC AGTGGTTTAA TCTATGTGCA
1951 TGGTTAGGGA GCCAGGCCTG GAATATTCAG TTTCCCTGCC CCTGTTAAAG
2001 AATCAGAGGT TGGGCAGTCA TCAAATTCAT CATAAAGACA TGGGCAAGTG
2051 TGTCTGTGGT TTCCAAGGCC CCCCTATGGA GAATCCAAAA GTATTTTCCA
2101 TTGCCGTGCT CTTTGAATGC AGACTTCTAT TTCCAGAAGT GACAGCACAA
2151 GTCTGAGTTG CTGTTTGGTC TGGTGACCTC AGACACACTA ATTTGAATTG
2201 AAAGCTAAGA GTAAAAATTT GCTGGTTACA GGCGAGTCAT ACTCTTGCAA
2251 GTAGTTAGCA AAGGGAGGCC CAAATTCTCA AGGTTGTTGA TGGGGAACTT
2301 GCCACTAAGA GAAGGCAGAG AGGTCCCTAG TGGGTATATT TGCTGCCAAG
2351 CCACTTGCCA AAGAAGAGGA ACCACAGAAA GAGAGACATC ATGACCAGGA
2401 GAAAAATGTG ACTAGACATG CTAACCTCCA GGTTTTTATA TATGACTTGA
2451 GTCTGCTGTA ATTGGCAGCA GAAATCCAAA TTTGTATGGT AGACCAAAAA
2501 GAACCAAATC CATAGGGTGA AATTTTGAGA CCTAGACTCT GTAAAAATAA
```

```
2551 TCCTAGTCTT CCTCCAGGGG TCAGTTCCTC ACAGTGGTTC TGTACCAAAA
2601 CTTGCCAAAT TCCTCCATGG CCAAGTGTTA AAATCTGTGT TTGGAAAATA
2651 GCGAATTAAC CTAAGACACA GAAGGCAGAC TGGGTGAGGA GACCTAGCAT
2701 GCCCTATTGG CAGTGCTCAG GAGCTGCATC CCACTTTTCC CTGCTCTGAA
2751 TCGAAGTCCT AGTTCCTTCC TTTGATTCTC CTTTGGTAGG TGGAATCAGT
2801 TAATGTTTTG AGAAACCTGC CTGGGCTCTG CCCTTAGTCA TGACATCTCG
2851 CTGAGCCAGA CCCACTCTGT TCCTTGGAAC CTAGAGCTGG AGTGAGGAGT
2901 AGAGGTCTCC GGCTATTCCA GAAAGAAAAG TGAGCCACAT GCAGGCTGAT
2951 GAATGCCGAC ACTTCCAGAA TGTATAGAAA TAGTCCCTGT CCTGGCCTGC
3001 CACTGACCCT GTCTGTATTT TCTCGGAGGT TGTTTTTCTC CTTCTCCTTC
3051 CCAGGAAGGT CTTTGTATGT CGAATCCAGT GCACTCAAGT TTGGCCAAGG
3101 GACTCCACAG CACCCAGAGG ACTGCATGCC TCAAGGTTTA TGTCACTCCT
3151 CTGCTGGGCT GTTCATTGTC ATTGCTGTGT TCAGGGACCT TTGGAAATAA
3201 AACCTGTTCT GTCCCAAATA AAACCAGCCT GTGATGTTCA AGGGACTGGA
3251 ATAAAGTGGC TTACGACCTG AAGGATTCTA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

Entry HS431350 from database EMBL:
human STS WI-15914.
Score = 1308, P = 3.1e-53, identities = 276/285

Entry HSG19929 from database EMBL:
human STS A002C26.
Score = 926, P = 1.5e-35, identities = 186/187

Entry AF052142 from database EMBL:
Homo sapiens clone 24665 mRNA sequence.
Score = 7378, P = 0.0e+00, identities = 1482/1487
3' UTR

Medline entries

93247712:
Neurocalcin family: a novel calcium-binding protein abundant in bovine central nervous system.

94045365:
Distinct regional localization of neurocalcin, a Ca(2+)-binding protein, in the bovine adrenal gland.

96407688:
Crystallization and preliminary X-ray crystallographic studies of recombinant bovine neurocalcin delta.

96066284:
Distribution pattern of three neural calcium-binding proteins (NCS-1, VILIP and recoverin) in chicken, bovine and rat retina.

Peptide information for frame 1

ORF from 121 bp to 699 bp; peptide length: 193
Category: strong similarity to known protein
Prosite motifs: EF_HAND (73-86)
EF_HAND (109-122)
EF_HAND (157-170)

```
  1 MGKQNSKLRP EVMQDLLEST DFTEHEIQEW YKGFLRDCPS GHLSMEEFKK
 51 IYGNFFPYGD ASKFAEHVFR TFDANGDGTI DFREFIIALS VTSRGKLEQK
101 LKWAFSMYDL DGNGYISKAE MLVIVQAIYK MVSSVMKMPE DESTPEKRTE
151 KIFRQMDTNR DGKLSLEEFI RGAKSDPSIV RLLQCDPSSA GQF
```

BLASTP hits

Entry JH0616 from database PIR:
neurocalcin (clone pCalN) - bovine

Score = 1001, P = 5.2e-101, identities = 192/193, positives = 192/193

Entry GGU91630_1 from database TREMBL:
product: "neurocalcin"; Gallus gallus neurocalcin mRNA, complete cds.
Score = 998, P = 1.1e-100, identities = 191/193, positives = 192/193

Entry NECD_BOVIN from database SWISSPROT:
NEUROCALCIN DELTA.
Score = 996, P = 1.8e-100, identities = 191/192, positives = 191/192

Entry S47565 from database PIR:
BDR-1 protein - human
Score = 934, P = 6.6e-94, identities = 174/193, positives = 187/193

Entry I50676 from database PIR:
gene Rem-1 protein - chicken >TREMBL:GGREM1_1 gene: "Rem-1"; G.gallus
rem-1 mRNA
Score = 933, P = 8.4e-94, identities = 174/193, positives = 186/193

Alert BLASTP hits for DKFZphfbr2_23b21, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23b21, frame 1

Report for DKFZphfbr2_23b21.1

[LENGTH] 193
[MW] 22215.30
[pI] 5.35
[HOMOL] PIR:JH0616 neurocalcin (clone pCalN) - bovine 1e-109
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YDR373w] 3e-54
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YKL190w] 2e-18
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YKL190w] 2e-18
[FUNCAT] 03.01 cell growth [S. cerevisiae, YKL190w] 2e-18
[FUNCAT] 13.04 homeostasis of other ions [S. cerevisiae, YKL190w] 2e-18
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL190w] 2e-18
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YBR109c] 0.001
[FUNCAT] 08.19 cellular import [S. cerevisiae, YBR109c] 0.001
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YBR109c] 0.001
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR109c] 0.001
[FUNCAT] 10.02.99 other morphogenetic activities [S. cerevisiae, YBR109c] 0.001
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YBR109c] 0.001
[BLOCKS] BL00018
[SCOP] d1rec__ 1.34.1.5.18 Recoverin [bovine (Bos taurus) 8e-55
[SCOP] d1jsa__ 1.34.1.5.17 Recoverin [human (Homo sapiens) 5e-58
[SCOP] d1tcob_ 1.34.1.5.16 Calcineurin regulatory subunit (B-chain 1e-06
[SCOP] d2mysc_ 1.34.1.5.15 Myosin Regulatory Chain [chicken (Gallu 2e-29
[SCOP] d1scmc_ 1.34.1.5.14 Myosin Regulatory Chain [bay scallo 5e-33
[SCOP] d2mysb_ 1.34.1.5.13 Myosin Essential Chain [chicken (Gallu 4e-26
[SCOP] d1scmb_ 1.34.1.5.12 Myosin Essential Chain [bay scallo 6e-27
[SCOP] d1clm__ 1.34.1.5.11 Calmodulin [Paramecium tetraurelia 1e-15
[SCOP] d4cln__ 1.34.1.5.10 Calmodulin [Drosophila melanogaster 2e-16
[SCOP] d1cfc__ 1.34.1.5.9 Calmodulin [African frog (Xenopus laevis) 2e-16
[SCOP] d1ahr__ 1.34.1.5.8 Calmodulin [chicken gallus gallus 4e-16
[SCOP] d3cln__ 1.34.1.5.7 Calmodulin [rat (Rattus rattus) 2e-16
[SCOP] d1trcb_ 1.34.1.5.6 Calmodulin [bovine (Bos taurus) 8e-08
[SCOP] d1cll__ 1.34.1.5.5 Calmodulin [human (Homo sapiens) 2e-16
[SCOP] d1rtpl_ 1.34.1.4.5 Parvalbumin [rat (Rattus rattus) 8e-06
[SCOP] d5tnc__ 1.34.1.5.2 Troponin C [turkey (Meleagris gallopavo) 3e-13
[SCOP] d1pvaa_ 1.34.1.4.3 Parvalbumin [pike (Esox lucius) 6e-06
[SCOP] d1tnp__ 1.34.1.5.1 Troponin C [chicken (Gallus gallus) 9e-11
[EC] 2.7.1.107 Diacylglycerol kinase 2e-08
[PIRKW] blocked amino end 1e-100
[PIRKW] phosphotransferase 2e-08
[PIRKW] duplication 4e-17
[PIRKW] tandem repeat 7e-06
[PIRKW] heterodimer 4e-17
[PIRKW] heart 6e-09
[PIRKW] zinc 2e-08
[PIRKW] serine/threonine-specific protein kinase 1e-06
[PIRKW] muscle contraction 1e-08
[PIRKW] acetylated amino end 4e-09
[PIRKW] ATP 2e-08
[PIRKW] skeletal muscle 6e-09

```
[PIRKW]        signal transduction 1e-91
[PIRKW]        protein kinase 2e-08
[PIRKW]        calcium binding 1e-100
[PIRKW]        alternative splicing 2e-13
[PIRKW]        methylated amino acid 1e-09
[PIRKW]        thin filaments 1e-08
[PIRKW]        lipoprotein 1e-101
[PIRKW]        cardiac muscle 6e-09
[PIRKW]        muscle 6e-09
[PIRKW]        myristylation 1e-100
[PIRKW]        EF hand 1e-101
[PIRKW]        retina 2e-51
[SUPFAM]       calcium-dependent protein kinase 2e-08
[SUPFAM]       unassigned calmodulin-related proteins 8e-41
[SUPFAM]       spec-related protein LpS1 7e-06
[SUPFAM]       calmodulin repeat homology 1e-101
[SUPFAM]       human diacylglycerol kinase 2e-08
[SUPFAM]       protein kinase C zinc-binding repeat homology 2e-08
[SUPFAM]       protein kinase homology 2e-08
[SUPFAM]       calmodulin 1e-101
[PROSITE]      EF_HAND 3
[PROSITE]      CK2_PHOSPHO_SITE     7
[PROSITE]      PKC_PHOSPHO_SITE     3
[PFAM]         EF hand
[KW]           All_Alpha
[KW]           3D

SEQ     MGKQNSKLRPEVMQDLLESTDFTEHEIQEWYKGFLRDCPSGHLSMEEFKKIYGNFFPYGD
1rec-   .......HHHHHHHHTTTTCCCHHHHHHHHHHHHHHHTTTTEEEHHHHHHHHHHHTTTTC

SEQ     ASKFAEHVFRTFDANGDGTIDFREFIIALSVTSRGKLEQKLKWAFSMYDLDGNGYISKAE
1rec-   HHHHHHHHHHHH-----CEEEHHHHHHHHHHHHHCCCGGGHHHHHHHHHHTTTTCCCEEHHH

SEQ     MLVIVQAIYKMVSSVMKMPEDESTPEKRTEKIFRQMDTNRDGKLSLEEFIRGAKSDPSIV
1rec-   HHHHHHHHHHCCTTGGGCTTTTTCHHHHHHHHHHHHCCTTTTEECHHHHHHHHHHCHHHH

SEQ     RLLQCDPSSAGQF
1rec-   HHHCCCH......
```

Prosite for DKFZphfbr2_23b21.1

PS00005	92->95	PKC_PHOSPHO_SITE	PDOC00005
PS00005	149->152	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00006	23->27	CK2_PHOSPHO_SITE	PDOC00006
PS00006	44->48	CK2_PHOSPHO_SITE	PDOC00006
PS00006	106->110	CK2_PHOSPHO_SITE	PDOC00006
PS00006	117->121	CK2_PHOSPHO_SITE	PDOC00006
PS00006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS00006	158->162	CK2_PHOSPHO_SITE	PDOC00006
PS00006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS00018	73->86	EF_HAND	PDOC00018
PS00018	109->122	EF_HAND	PDOC00018
PS00018	157->170	EF_HAND	PDOC00018

Pfam for DKFZphfbr2_23b21.1

```
HMM_NAME      EF hand

HMM                *MFrmMDkDGDGyIDFEEFmeMMkem*
                    +FR +D +GDG+IDF EF+ +++
Query          68  VFRTFDANGDGTIDFREFIIALSVT     92

30.75    100   128    1    29 dkfzphfbr2_23b21.1 nearly identical to bovine neurocalcin
  Alignment to HMM consensus:
Query              *EIqEMFrmMDkDGDGyIDFEEFmeMMkem*
                     ++++F+M+D DG+GYI++ E++++++++
  dkfzphfbr2   100  KLKWAFSMYDLDGNGYISKAEMLVIVQAI     128

Query          176    1    29 dkfzphfbr2_23b21.1 nearly identical to bovine neurocalcin
  Alignment to HMM consensus:
HMM                *EIqEMFrmMDkDGDGyIDFEEFmeMMkem*
                     +++FR MD+++DG+++ EEF++  K+
Query         148  RTEKIFRQMDTNRDGKLSLEEFIRGAKSD     176
```

DKFZphfbr2_23f2
----

group: brain derived

DKFZphfbr2_23f2 encodes a novel 182 amino acid protein with weak similarity to S. pombe Vps29p.

No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to Vps29p

complete cDNA, complete cds, EST hits
S.cerevisiae and S.pombe Vps29p are involved in vacuolar protein sorting
part of the cDNA is encoded by HSAC2350, splice pattern 4 exons

Sequenced by AGOWA

Locus: /map="12q24"

Insert length: 1016 bp
Poly A stretch at pos. 996, polyadenylation signal at pos. 974

```
   1 GAATGGGGAG GAGCCAGAGG AAGAGGGCGG CGACGGTGGT GGTGACTGAG
  51 CGGAGCCCGG TGACAGGATG TTGGTGTTGG TATTAGGAGA TCTGCACATC
 101 CCACACCGGT GCAACAGTTT GCCAGCTAAA TTCAAAAAAC TCCTGGTGCC
 151 AGGAAAAAATT CAGCACATTC TCTGCACAGG AAACCTTTGC ACCAAAGAGA
 201 GTTATGACTA CCTCAAGACT CTGGCTGGTG ATGTTCATAT TGTGAGAGGA
 251 GACTTCGATG AGAATCTGAA TTATCCAGAA CAGAAAGTTG TGACTGTTGG
 301 ACAGTTCAAA ATTGGTCTGA TCCATGGACA TCAAGTTATT CCATGGGGAG
 351 ATATGGCCAG CTTAGCCCTG TTGCAGAGGC AATTTGATGT GGACATTCTT
 401 ATCTCGGGAC ACACACACAA ATCTGAAGCA TTTGAGCATG AAAATAAATT
 451 CTACATTAAT CCAGGTTCTG CCACTGGGGC ATATAATGCC TTGGAAACAA
 501 ACATTATTCC ATCATTTGTG TTGATGGATA TCCAGGCTTC TACAGTGGTC
 551 ACCTATGTGT ATCAGCTAAT TGGAGATGAT GTGAAAGTAG AACGAATCGA
 601 ATACAAAAAA CCTTAAAGCC AGGCCTGTCT TGATGATTTT TGGTTTTTTT
 651 TCATTGTCCT GTTGAAATCA AGTAATTAAA CATTTAAGAG CCACAAAATT
 701 GTATCACTTT TATAATATTT TGCAGTAAAA TATAATACCA TCTTCTCTGT
 751 TAATACATAA TTGCTCCAAG CTTCCTGTAA ACTATAAGAA TATATTTAGT
 801 TTACAGTATA TGGATTCTAT GAAAAAATGT CCACAACACA GTAATTGGTC
 851 ACTTGTTAAG AAAAATTTAT CCTTGTAAGT ATCTTCAAAG TTGATATTTG
 901 GAACTTTATT CCAAAAGTAG TGCATGTGGA GAAAGAATCT AGACTTTCTT
 951 GTATACATTT TTCTCTTCTC CAGTAATAAA CAATTACCTT TCATTGAAAA
1001 AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HSAC2350 from database EMBLNEW:
Homo sapiens 12q24 PAC P424M6 Length = 167,217

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 68 bp to 613 bp; peptide length: 182
Category: similarity to known protein
Prosite motifs: RGD (60-63)

```
  1 MLVLVLGDLH IPHRCNSLPA KFKKLLVPGK IQHILCTGNL CTKESYDYLK
 51 TLAGDVHIVR GDFDENLNYP EQKVVTVGQF KIGLIHGHQV IPWGDMASLA
101 LLQRQFDVDI LISGHTHKSE AFEHENKFYI NPGSATGAYN ALETNIIPSF
```

151 VLMDIQASTV VTYVYQLIGD DVKVERIEYK KP

BLASTP hits

Entry CEZK1128_6 from database TREMBL:
"ZK1128.1"; Caenorhabditis elegans cosmid ZK1128
Length = 523
Score = 400 (140.8 bits), Expect = 2.3e-37, P = 2.3e-37
Identities = 81/150 (54%), Positives = 106/150 (70%)

Entry S46793 from database PIR:
hypothetical protein YHR012c - yeast (Saccharomyces cerevisiae)
Length = 282
Score = 180 (63.4 bits), Expect = 3.7e-37, Sum P(3) = 3.7e-37
Identities = 35/71 (49%), Positives = 44/71 (61%)

Entry AB011824_1 from database TREMBL:
"Vps29"; Schizosaccharomyces pombe mRNA for Vps29,
partial cds. Schizosaccharomyces pombe (fission yeast)
Length = 176
Score = 189 (66.5 bits), Expect = 2.7e-27, Sum P(2) = 2.7e-27
Identities = 33/72 (45%), Positives = 50/72 (69%)

Alert BLASTP hits for DKFZphfbr2_23f2, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23f2, frame 2

Report for DKFZphfbr2_23f2.2

[LENGTH] 182
[MW] 20445.84
[pI] 6.29
[HOMOL] TREMBL:CEZK1128_6 gene: "ZK1128.8"; Caenorhabditis elegans cosmid ZK1128 2e-51
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YHR012w] 1e-27
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YHR012w] 1e-27
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YHR012w] 1e-27
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YHR012w] 1e-27
[FUNCAT] 09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YHR012w] 1e-27
[FUNCAT] r general function prediction [M. jannaschii, MJ0623] 1e-16
[BLOCKS] BL01269D
[BLOCKS] BL01269A
[PROSITE] RGD 1
[PROSITE] MYRISTYL 4
[PROSITE] PKC_PHOSPHO_SITE 1
[KW] Alpha_Beta

SEQ MLVLVLGDLHIPHRCNSLPAKFKKLLVPGKIQHILCTGNLCTKESYDYLKTLAGDVHIVR
PRD ccceeecccccccccccchhhhhhhhhcceeeeeecccccchhhhhhhhhhhceeeee

SEQ GDFDENLNYPEQKVVTVGQFKIGLIHGHQVIPWGDMASLALLQRQFDVDILISGHTHKSE
PRD ccccccccccceeeeecccceeeeecccccccccchhhhhhhhhhcceeeeeccccccc

SEQ AFEHENKFYINPGSATGAYNALETNIIPSFVLMDIQASTVVTYVYQLIGDDVKVERIEYK
PRD ccccccccccccccccccccccccccccceeeeeccccceeeeeeeeeccccceeeeeeeec

SEQ KP
PRD cc

Prosite for DKFZphfbr2_23f2.2

PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	83->89	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	137->143	MYRISTYL	PDOC00008
PS00016	60->63	RGD	PDOC00016

(No Pfam data available for DKFZphfbr2_23f2.2)

DKFZphfbr2_23124
----

group: intracellular transport and trafficking

DKFZphfbr2_23124.2 encodes a novel 348 amino acid protein with similarity to human glycoprotein gp36b and canine VIP36 glycoprotein.

The vesicular protein VIP36 (36 kDa vesicular integral membrane protein) shows homology to leguminous plant lectins. The protein is localized to the Golgi apparatus, endosomal and vesicular structures and the plasma membrane. VIP36 binds to sugar residues of glycosphingolipids and/or glycosylphosphatidyl-inositol anchors and might provide a link between the extracellular/luminal face of glycolipid rafts and the cytoplasmic protein segregation machinery. Gp36 is located within the endoplasmatic reticulum. For the novel protein, a lectin character is predicted. Due to the intracellular localisation of the homolog proteins, it should be involved in intracellular transport and trafficking.

The new protein can find application in modulating/blocking intracellular transport and trafficking.

strong similarity to human GP36b glycoprotein

complete cDNA, complete cds, EST hits
potential start at Bp 29 matches kozak consensua ANNatgG
similarity to lectins,

Sequenced by AGOWA

Locus: /map="2"

Insert length: 2416 bp
Poly A stretch at pos. 2394, no polyadenylation signal found

```
   1 GGGGGATGAA GGGTCGTTGG TGGGAAAGAT GGCGGCGACT CTGGGACCCC
  51 TTGGGTCGTG GCAGCAGTGG CGGCGATGTT TGTCGGCTCG GGATGGGTCC
 101 AGGATGTTAC TCCTTCTTCT TTTGTTGGGG TCTGGGCAGG GGCCACAGCA
 151 AGTCGGGGCG GGTCAAACGT TCGAGTACTT GAAACGGGAG CACTCGCTGT
 201 CGAAGCCCTA CCAGGGTGTG GGCACAGGCA GTTCCTCACT GTGGAATCTG
 251 ATGGGCAATG CCATGGTGAT GACCCAGTAT ATCCGCCTTA CCCCAGATAT
 301 GCAAAGTAAA CAGGGTGCCT TGTGGAACCG GGTGCCATGT TTCCTGAGAG
 351 ACTGGGAGTT GCAGGTGCAC TTCAAAATCC ATGGACAAGG AAAGAAGAAT
 401 CTGCATGGGG ATGGCTTGGC AATCTGGTAC ACAAAGGATC GGATGCAGCC
 451 AGGGCCTGTG TTTGGAAACA TGGACAAATT TGTGGGGCTG GGAGTATTTG
 501 TAGACACCTA CCCCAATGAG GAGAAGCAGC AAGAGCGGGT ATTCCCCTAC
 551 ATCTCAGCCA TGGTGAACAA CGGCTCCCTC AGCTATGATC ATGAGCGGGA
 601 TGGGCGGCCT ACAGAGCTGG GAGGCTGCAC AGCCATTGTC CGCAATCTTC
 651 ATTACGACAC CTTCCTGGTG ATTCGCTACG TCAAGAGGCA TTTGACGATA
 701 ATGATGGATA TTGATGGCAA GCATGAGTGG AGGGACTGCA TTGAAGTGCC
 751 CGGAGTCCGC CTGCCCCGCG GCTACTACTT CGGCACCTCC TCCATCACTG
 801 GGGATCTCTC AGATAATCAT GATGTCATTT CCTTGAAGTT GTTTGAACTG
 851 ACAGTGGAGA GAACCCCAGA AGAGGAAAAG CTCCATCGAG ATGTGTTCTT
 901 GCCCTCAGTG GACAATATGA AGCTGCCTGA GATGACAGCT CCACTGCCGC
 951 CCCTGAGTGG CCTGGCCCTC TTCCTCATCG TCTTTTTCTC CCTGGTGTTT
1001 TCTGTATTTG CCATAGTCAT TGGTATCATA CTCTACAACA AATGGCAGGA
1051 ACAGAGCCGA AAGCGCTTCT ACTGAGCCCT CCTGCTGCCA CCACTTTTGT
1101 GACTGTCACC CATGAGGTAT GGAAGGAGCG GGCACTGGCC TGAGCATGCA
1151 GCCTGGAGAG TGTTCTTGTC TCTAGCAGCT GGTTGGGGAC TATATTCTGT
1201 CACTGGAGTT TTGAATGCAG GGACCCCGCA TTCCCATGGT TGTGCATGGG
1251 GACATCTAAC TCTGGTCTGG GAAGCCACCC ACCCCAGGGC AATGCTGCTG
1301 TGATGTGCCT TTCCCTGCAG TCCTTCCATG TGGGAGCAGA GGTGTGAAGA
1351 GAATTTACGT GGTTGTGATG CCAAAATCAC GGAACAGAAT TTCATAGCCC
1401 AGGCTGCCGT GTTGTTTGAC TCAGAAGGCC CTTCTACTTC AGTTTTGAAT
1451 CCACAAAGAA TTAAAAACTG GTAACACCAC AGGCTTTCTG ACCATCCATT
1501 CGTTGGGTTT TGCATTTGAC CCAACCCTCT GCCTACCTGA GGAGCTTTCT
1551 TTGGAAACCA GGATGGAAAC TTCTTCCCTG CCTTACCTTC CTTTCACTCC
1601 ATTCATTGTC CTCTCTGTGT GCAACCTGAG CTGGGAAAGG CATTTGGATG
1651 CCTCTCTGTT GGGGCCTGGG GCTGCAGAAC ACACCTGCGT TTCGCTGGCC
1701 TTCATTAGGT GGCCCTAGGG AGATGGCTTT CTGCTTTGGA TCACTGTTCC
1751 CTAGCATGGG TCTTGGGTCT ATTGGCATGT CCATGGCCTT CCCAATCAAG
1801 TCTCTTCAGG CCCTCAGTGA AGTTTGGCTA AAGGTTGGTG TAAAAATCAA
1851 GAGAAGCCTG GAAGACACCA TGGATGCCAT GGATTAGCTG TGCAACTGAC
1901 CAGCTCCAGG TTTGATCAAA CCAAAAGCAA CATTTGTCAT GTGGTCTGAC
1951 CATGTGGAGA TGTTTCTGGA CTTGCTAGAG CCTGCTTAGC TGCATGTTTT
2001 GTAGTTACGA TTTTTGGAAT CCCTCTTTGA GTGCTGAAAG TGTAAGGAAG
2051 CTTTCTTCTT ACACCTTGGG CTTGGATATT GCCCAGAGAA GAAATTTGGC
2101 TTTTTTTTCT TAATGGACAA GGGACAGTTG CTGTTCTCAT GTTCCAAGTC
2151 TGAGAGCAAC AGACCCTCAT CATCTGTGCC TGGAAGAGTT CACTGTCATT
2201 GAGCAGCACA GCCTGAGTGC TGGCCTCTGT CAACCCTTAT TCCACTGCCT
```

```
2251 TATTTGACAA GGGGTTACAT GCTGCTCACC TTACTGCCCT GGGATTAAAT
2301 CAGTTACAGG CCAGAGTCTC CTTGGAGGGC CTGGAACTCT GAGTCCTCCT
2351 ATGAACCTCT GTAGCCTAAA TGAAATTCTT AAAATCACCG ATGGAACCAA
2401 AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HS622145 from database EMBL:
human STS WI-6746.
Score = 1079, P = 5.1e-43, identities = 219/223

Entry G42541 from database EMBLNEW:
SHGC-58649 Human Homo sapiens STS genomic, sequence tagged site.
Score = 1091, P = 1.7e-43, identities = 219/220

Medline entries

94265253:
A putative novel class of animal lectins in the secretory pathway homologous to leguminous
lectins.

94208543:
VIP36, a novel component of glycolipid rafts and exocytic carrier vesicles in epithelial cells.

Peptide information for frame 2

ORF from 29 bp to 1072 bp; peptide length: 348
Category: strong similarity to known protein

```
  1 MAATLGPLGS WQQWRRCLSA RDGSRMLLLL LLLGSGQGPQ QVGAGQTFEY
 51 LKREHSLSKP YQGVGTGSSS LWNLMGNAMV MTQYIRLTPD MQSKQGALWN
101 RVPCFLRDWE LQVHFKIHGQ GKKNLHGDGL AIWYTKDRMQ PGPVFGNMDK
151 FVGLGVFVDT YPNEEKQQER VFPYISAMVN NGSLSYDHER DGRPTELGGC
201 TAIVRNLHYD TFLVIRYVKR HLTIMMDIDG KHEWRDCIEV PGVRLPRGYY
251 FGTSSITGDL SDNHDVISLK LFELTVERTP EEEKLHRDVF LPSVDNMKLP
301 EMTAPLPPLS GLALFLIVFF SLVFSVFAIV IGIILYNKWQ EQSRKRFY
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23124, frame 2

PIR:G01447 GP36b glycoprotein - human, N = 1, Score = 1001, P = 5.9e-101

SWISSPROT:VP36_CANFA VESICULAR INTEGRAL-MEMBRANE PROTEIN VIP36 PRECURSOR (VIP36)., N = 1, Score = 990, P = 8.6e-100

TREMBL:CET04G9_2 gene: "T04G9.3"; Caenorhabditis elegans cosmid T04G9., N = 1, Score = 614, P = 6e-60

PIR:S42626 ER-golgi intermediate compartment protein - human, N = 2, Score = 397, P = 1e-42

>PIR:G01447 GP36b glycoprotein - human
Length = 356

HSPs:

Score = 1001 (150.2 bits), Expect = 5.9e-101, P = 5.9e-101
Identities = 197/356 (55%), Positives = 256/356 (71%)

```
Query:    1 MAATLGPLGSWQQWRRCLSARDG--------SRMLLLLLLLGSGQGPQQVGAGQTFEYLK 52
              MAA  G +  W    RRCL  R G         + L LLLLGS       + G + E+LK
Sbjct:    1 MAAE-GWIWRWGWGRRCLG-RPGLLGPGPGPTTPLFLLLLLGSVTA--DITDGNS-EHLK 55
```

```
Query:   53 REHSLSKPYQGVGTGSSSLWNLMGNAMVMTQYIRLTPDMQSKQGALWNRVPCFLRDWELQ 112
            REHSL KPYQGVG+ S  LW+  G+ M+ +QY+RLTPD +SK+G++WN  PCFL+DWE+
Sbjct:   56 REHSLIKPYQGVGSSSMPLWDFQGSTMLTSQYVRLTPDERSKEGSIWNHQPCFLKDWEMH 115

Query:  113 VHFKIHGQGKKNLHGDGLAIWYTKDRMQPGPVFGNMDKFVGLGVFVDTYPNEEKQQERVF 172
            VHFK+HG GKKNLHGDG+A+WYT+DR+ PGPVFG+ D F GL +F+DTYPN+E   ERVF
Sbjct:  116 VHFKVHGTGKKNLHGDGIALWYTRDRLVPGPVFGSKDNFHGLAIFLDTYPNDETT-ERVF 174

Query:  173 PYISAMVNNGSLSYDHERDGRPTELGGCTAIVRNLHYDTFLVIRYVKRHLTIMMDIDGKH 232
            PYIS MVNNGSLSYDH +DGR TEL GCTA  RN  +DTFL +RY +  LT+M D++ K+
Sbjct:  175 PYISVMVNNGSLSYDHSKDGRWTELAGCTADFRNRDHDTFLAVRYSRGRLTVMTDLEDKN 234

Query:  233 EWRDCIEVPGVRLPRGYYFGTSSITGDLSDNHDVISLKLFELTVERTPEEEKLHRDVFLP 292
            EW++CI++ GVRLP GYYFG S+ TGDLSDNHD+IS+KLF+L VE TP+EE +       P
Sbjct:  235 EWKNCIDITGVRLPTGYYFGASAGTGDLSDNHDIISMKLFQLMVEHTPDEESIDWTKIEP 294

Query:  293 SVDNMKLPEMTAPLP-------PLSGLALFLIVFFSLVFSVFAIVIGIILYNKWQEQSRK 345
            SV+ +K P+     P       PL+G  +FL++  +L+  V   V+G +++ K QE++ K
Sbjct:  295 SVNFLKSPKDNVDDPTGNFRSGPLTGWRVFLLLLCALLGIVVCAVVGAVVFQKRQERN-K 353

Query:  346 RFY 348
            RFY
Sbjct:  354 RFY 356
```

Pedant information for DKFZphfbr2_23124, frame 2

Report for DKFZphfbr2_23124.2

```
[LENGTH]     348
[MW]         39711.10
[pI]         8.55
[HOMOL]      PIR:G01447 GP36b glycoprotein - human 1e-101
[PIRKW]      lectin 2e-37
[PIRKW]      transmembrane protein 2e-37
[PIRKW]      endoplasmic reticulum 2e-37
[PIRKW]      Golgi apparatus 2e-37
[PROSITE]    AMIDATION        1
[PROSITE]    MYRISTYL         5
[PROSITE]    CK2_PHOSPHO_SITE     2
[PROSITE]    GLYCOSAMINOGLYCAN    1
[PROSITE]    PKC_PHOSPHO_SITE     3
[PROSITE]    ASN_GLYCOSYLATION    1
[KW]         Alpha_Beta
[KW]         SIGNAL_PEPTIDE 39
[KW]         LOW_COMPLEXITY     7.76 %

SEQ    MAATLGPLGSWQQWRRCLSARDGSRMLLLLLLLLGSGQGPQQVGAGQTFEYLKREHSLSKP
SEG    ..........................xxxxxxx............................
PRD    cccccccccccccccccccccccchhhhhhhhhhhcccccccccccchhhhhhhhhhhhccc

SEQ    YQGVGTGSSSLWNLMGNAMVMTQYIRLTPDMQSKQGALWNRVPCFLRDWELQVHFKIHGQ
SEG    ............................................................
PRD    cccccccccceeeccccccccccceeeecccchhhhhccccccccccchhhhhhhheeeeecc

SEQ    GKKNLHGDGLAIWYTKDRMQPGPVFGNMDKFVGLGVFVDTYPNEEKQQERVFPYISAMVN
SEG    ............................................................
PRD    cccccccceeeeeeccccccccccccccccccceeeeeeccccccccccccccceeeeeec

SEQ    NGSLSYDHERDGRPTELGGCTAIVRNLHYDTFLVIRYVKRHLTIMMDIDGKHEWRDCIEV
SEG    ............................................................
PRD    ccccccccccccccccccccccccccccccccceeeehhhhhheeeeecccccccccccc

SEQ    PGVRLPRGYYFGTSSITGDLSDNHDVISLKLFELTVERTPEEEKLHRDVFLPSVDNMKLP
SEG    ............................................................
PRD    ccccccccccccccccccccccccchhhhhhhhhhhhcccccccccccccccccccccc

SEQ    EMTAPLPPLSGLALFLIVFFSLVFSVFAIVIGIILYNKWQEQSRKRFY
SEG    ...........xxxxxxxxxxxxxxxxxxxx.................
PRD    ccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccc
```

Prosite for DKFZphfbr2_23124.2

PS00001	181->185	ASN_GLYCOSYLATION	PDOC00001
PS00002	35->39	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	19->22	PKC_PHOSPHO_SITE	PDOC00005

PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS00006	279->283	CK2_PHOSPHO_SITE	PDOC00006
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	63->69	MYRISTYL	PDOC00008
PS00008	65->71	MYRISTYL	PDOC00008
PS00008	96->102	MYRISTYL	PDOC00008
PS00008	198->204	MYRISTYL	PDOC00008
PS00009	120->124	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_23124.2)

DKFZphfbr2_23n16
----

group: signal transduction

DKFZphfbr2_23n16.1 encodes a novel 292 amino acid protein with weak similarity to putative phosphatidylinositol-4-phosphate 5-kinase of Arabidopsis thaliana.

The novel proteins contains a WW domain which has been originally described as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes. Examples of proteins containing the WW domain are Dystrophin, Utrophin, vertebrate YAP protein (binds the SH3 domain of the Yes oncoprotein), murine NEDD-4 (embryonic development and differentiation of the central nervous system), IQGAP (human GTPase activating protein acting on ras). Therefore the new protein should be involved in intracellular signal transduction.

The new protein can find application in modulating/blocking intracellular signal transduction pathways.

similarity to putative phosphatidylinositol-4-phosphate 5-kinase

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 2936 bp
Poly A stretch at pos. 2916, polyadenylation signal at pos. 2873

```
   1 GGGGGCGCTC CCGAGAAAGA GTGAGGGCGC GACGCGCACC AACGGTGGAG
  51 GGATGTTTCA GCAGCCCCTG AGAAGGAAGA GGAGGAAGCT GAGGGCCCGC
 101 TGAGGGCGCA GGACCTGAGG GAGTCCTACA TCCAGCTCGT CCAGGGTGTG
 151 CAGGAGTGGC AGGATGGTTG CATGTACCAG GGGGAGTTTG GGTTGAACAT
 201 GAAGCTTGGA TATGGCAAAT TCTCTTGGCC CACAGGCGAG TCATACCATG
 251 GGCAGTTTTA CCGGGACCAC TGCCATGGCC TGGGTACCTA CATGTGGCCA
 301 GATGGCTCCA GTTTCACGGG CACATTTTAC CTCAGCCACC GAGAAGGCTA
 351 CGGCACCATG TACATGAAGA CACGGCTTTT CCAGACTCAC TGCCACAACG
 401 ACATTGTCAA CCTTCTCCTG GACTGTGGGG CCGACGTGAA CAAGTGCTCA
 451 GATGAGGGTC TCACGGCACT CAGCATGTGT TTCCTCCTCC ACTACCCCGC
 501 CCAGTCCTTC AAGCCCAATG TTGCTGAACG GACCATACCT GAGCCCCAGG
 551 AACCTCCAAA ATTCCCAGTT GTTCCAATCC TTTCATCATC ATTTATGGAC
 601 ACAAACCTGG AGTCTCTGTA CTATGAGGTG AACGTGCCTT CCCAGGGTAG
 651 CTATGAGCTG AGGCCACCGC CAGCACCACT GCTCCTGCCA CGCGTCTCAG
 701 GCAGCCACGA GGGCGGCCAC TTCCAGGACA CCGGGCAGTG TGGGGGGTCC
 751 ATAGACCACA GGAGCAGCTC TCTGAAGGGG GACTCCCCGT TGGTGAAGGG
 801 CAGCCTTGGC CATGTGGAAA GCGGGCTTGA GGACGTGTTG GGAGACACAG
 851 ACCGGGGCAG TCTGTGCAGT GCTGAGACGA AATTTGAGTC CAACTTGTGT
 901 GTGTGCGACT TCTCCATCGA GCTCTCGCAG GCCATGCTGG AGAGAAGCGC
 951 CCAGTCCCAC AGCTTGCTGA AGATGGCCTC GCCCTCACCG TGCACCAGCA
1001 GCTTCGACAA AGGGACCATG CGGAGGATGG CGCTGTCCAT GATCGAGTAG
1051 GTCCTGGCAC CAGCTGGTGG GGGTGGAGGG CCACCATCAG GGCTGAATCC
1101 TATGCTCAGC AGACCCACGT CTCTTCCCTG TGCCAGTGGG AGGCGTTGTG
1151 TCTGGAGATG TGTGTCTGAA TGTGTGAGCA TCCCTGTGTC GGTGGCTCCA
1201 TGCCATGGCC AGCCCTGTGG GGGTGCCACG GTGACGGGCT GTTTTCAGTG
1251 CCACCCCAGC CCTGTGGGGG TGCCACGGTG ACGGGCTGTT TTCAGTACCA
1301 CGCCAGCCCT GCTTTGGCCT TTGGCACTGG CCTGAAGTGT CTCTGTGGGA
1351 GCCTCAGCAG GGGCCACTGT CAGGGGTCCT ATCCTAGCCA TAGTGCACGT
1401 GAGTGACACC TGCCTGGGCA GCTCTCACAC CCCTGCTGTC CACCCTGTCT
1451 ATACCAGTGT GTCTCAAAAT GTGGTCTATG CACCCCCGGG GGTCCAAGAC
1501 CCTTTCAGGG AGTCTGTGGG GTCAAAATGA TTCTCTTGAT AACCCTGAGA
1551 CTCTGTTAGC CTTCTCCTTG TGTTGATGTT GGTGGATGGT ATGAAGACAG
1601 GGCCGTGCAG ACCACCAGCC CCCAGCGTGC AGGGCAGCAG TGCCCGGCCT
1651 GCTTGGGGGC ATGGTATTCC TTCACCACGG TGTGCACTTG CGGGGATGCC
1701 TGTCTCACTG AAGAATGCCT TTGACTAAGC AGAAAAGCAA TGACAAATTG
1751 CATTAAATCT TGCTCCTTGC GTACACACCC CTCGAATATT CTGGGTCGGA
1801 AAACATGGGA AGGACACTGA TGTGTGTCTG CCACAGACCA AGGCACACCG
1851 CTTCCCCGCA AGAAGCGCTT CCCCCAGGGC CAGAGTAGCA ACAGAATGCG
1901 GCATCTTCCC AACCTCCTGC CCCATTTTTG ATTGGAAGAA TGACCACTGG
1951 TATGTGGCTG TTCATTCTCC TGAACACAGC CTGCCACTTT AAGGAAAAACA
2001 TATGACACTA TTTGTTGCTG GCGAAATTTA CATTTTCAAG TGAATAGCAG
2051 AATTCTGGAC ACTTGCCACC ACCACCAAAA CCTTCATAGC TTCCCTTAAC
2101 TTTGAGACAT GGGTGTTCAG AGGTTTTTCA CGTGAGATGG CGTTAGCAGC
2151 GCAGTTTTGT GATACTGCCT GAAGACATGC CGACAGTGCC CAGATCTCTT
```

```
2201 CTATTGGTGA GCCAGCTTTT CCCACACGGC CAAGTTCTGA TGTTGAACCA
2251 TTGCCAGGTG GGTGAAGATC CATTGACAGT GAGAGGTGGG CCCGTGGGCT
2301 TCAGTGCAGC CAGGCGCAGA AGGCTGGTTC ATGAGTGTCC AGCTCCGCCA
2351 GGTAGCTAGC TCACCACCCC CAGCCTGGGT TCATGTAGTT CAAATAGGAA
2401 GACCACGATG ATCAGAAAGG CTGCTCAAAT ACTCCTTCGT CCAGCCGCGT
2451 ACCTGGGGGA GGCTGAATCT CCACTCACTT CCACCAAGGC TGTGCAGAGC
2501 AGATAGGGGA ATCCAGCAAA GGTGGAAAAC AGTGCCATCC TTCTCCCCAA
2551 CTGGTTTTGT TTTGTAAAAT AACTTTTTGT GACAGTGTTA CTTATTAGTA
2601 ACATGCAGTG GGTTTGTTAT GGTTAACAAG TTGGTGAGCA TTATTGAGAG
2651 GTGAAGCCAG CTGAGCTTCT GGGTTGGGTG GGGACTTGGA GAACTTTTGT
2701 GTCTAGCTAA AGGATTGTAA ATGCACCAAT CAATGCTCAG TGTCTAGCTA
2751 AAGGATTGTA AATGCACCAA TCAGCACTCT GTAAAATTGA CCAATCAGCG
2801 TTCTGTAAAA TGGACCAATC AGTGGTCTGT AAAATGGACC AGTCAGCAGG
2851 ATGTGGGCGG GGCCAAAAAA GGGAATAAAA GCTGGCCACC GCCAGGCTCC
2901 CCACCAGCCT GCAGCGAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 172 bp to 1047 bp; peptide length: 292
Category: similarity to unknown protein
Prosite motifs: WW_DOMAIN_1 (19-24)

```
  1 MYQGEFGLNM KLGYGKFSWP TGESYHGQFY RDHCHGLGTY MWPDGSSFTG
 51 TFYLSHREGY GTMYMKTRLF QTHCHNDIVN LLLDCGADVN KCSDEGLTAL
101 SMCFLLHYPA QSFKPNVAER TIPEPQEPPK FPVVPILSSS FMDTNLESLY
151 YEVNVPSQGS YELRPPPAPL LLPRVSGSHE GGHFQDTGQC GGSIDHRSSS
201 LKGDSPLVKG SLGHVESGLE DVLGDTDRGS LCSAETKFES NLCVCDFSIE
251 LSQAMLERSA QSHSLLKMAS PSPCTSSFDK GTMRRMALSM IE
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23n16, frame 1

TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for AtPIP5K1, complete cds., N = 2, Score = 138, P = 1.1e-06

TREMBL:AF019380_1 product: "putative phosphatidylinositol-4-phosphate 5-kinase"; Arabidopsis thaliana putative phosphatidylinositol-4-phosphate 5-kinase mRNA, complete cds., N = 2, Score = 138, P = 1.4e-06

PIR:T02098 probable phosphatidylinositol-4-phosphate 5-kinase - Arabidopsis thaliana, N = 2, Score = 135, P = 6.7e-06

```
>TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for
     AtPIP5K1, complete cds.
          Length = 683

 HSPs:

 Score = 138 (20.7 bits), Expect = 1.1e-06, Sum P(2) = 1.1e-06
 Identities = 23/61 (37%), Positives = 35/61 (57%)

Query:     1 MYQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFYLSHREGY 60
             MY+G++         G GKFSWP+G +Y G+F        G GT+     DG ++ GT+      + G+
Sbjct:    34 MYEGDWKRGKASGKGKFSWPSGATYEGEFKSGRMEGFGTFTGADGDTYRGTWVADRKHGH 93

Query:    61 G 61
             G
Sbjct:    94 G 94
```

```
 Score = 112 (16.8 bits), Expect = 9.7e-04, Sum P(2) = 9.7e-04
 Identities = 19/51 (37%), Positives = 27/51 (52%)

Query:     12 LGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFYLSHREGYGT 62
              +G GK+ W  G  Y G + R     G G + WP G+++ G F     EG+GT
Sbjct:     22 IGSGKYLWKDGCMYEGDWKRGKASGKGKFSWPSGATYEGEFKSGRMEGFGT 72

 Score = 97 (14.6 bits), Expect = 4.4e-02, Sum P(2) = 4.3e-02
 Identities = 19/60 (31%), Positives = 32/60 (53%)

Query:      2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFYLSHREGYG 61
              Y+GEF       G+G F+   G++Y G +  D  HG G   + +G  + GT+  + ++G G
Sbjct:     58 YEGEFKSGRMEGFGTFTGADGDTYRGTWVADRKHGHGQKRYANGDFYEGTWRRNLQDGRG 117

 Score = 93 (14.0 bits), Expect = 1.2e-01, Sum P(2) = 1.1e-01
 Identities = 18/62 (29%), Positives = 34/62 (54%)

Query:      2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFYLSHREGYG 61
              Y+G +  + K G+G+  +  G+ Y G + R+  G G Y+W +G+ +TG + +    G G
Sbjct:     81 YRGTWVADRKHGHGQKRYANGDFYEGTWRRNLQDGRGRYVWRNGNQYTGEWRIGVISGKG 140

Query:     62 TM 63
               +
Sbjct:    141 LL 142

 Score = 91 (13.7 bits), Expect = 2.0e-01, Sum P(2) = 1.8e-01
 Identities = 18/51 (35%), Positives = 24/51 (47%)

Query:      2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTF 52
              Y GE+ + +  G G   WP G  Y G +     G G + W DGSS  G +
Sbjct:    127 YTGEWRIGVISGKGLLVWPNGNRYEGLWENGIPKGNGVFTWSDGSSCVGAW 177

 Score = 90 (13.5 bits), Expect = 2.6e-01, Sum P(2) = 2.3e-01
 Identities = 17/60 (28%), Positives = 31/60 (51%)

Query:      2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFYLSHREGYG 61
              Y+G +  N++ G G++ W  G  Y G++      G G  +WP+G+ + G +     +G G
Sbjct:    104 YEGTWRRNLQDGRGRYVWRNGNQYTGEWRIGVISGKGLLVWPNGNRYEGLWENGIPKGNG 163

 Score = 45 (6.8 bits), Expect = 1.1e-06, Sum P(2) = 1.1e-06
 Identities = 14/62 (22%), Positives = 26/62 (41%)

Query:    215 VESGLEDVLGDTDRGSLCSAETKFESNLCVCDF--SIELSQAMLERSAQSHSLLKMASPS 272
              V+SG   + G+    +C  E+  E+    CD  ++E S   +R +       +    +
Sbjct:    205 VDSGAGSLGGEKVFPRICIWESDGEAGDITCDIIDNVEASMIYRDRISVDRDGFRQFKKN 264

Query:    273 PC 274
              PC
Sbjct:    265 PC 266
```

Pedant information for DKFZphfbr2_23n16, frame 1

Report for DKFZphfbr2_23n16.1

```
[LENGTH]        292
[MW]            32214.44
[pI]            5.51
[HOMOL]         TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for AtPIP5K1,
complete cds. 7e-08
[BLOCKS]        BL01137A Hypothetical YBL055c/yjjV family proteins
[PROSITE]       WW_DOMAIN_1      1
[PROSITE]       MYRISTYL         5
[PROSITE]       CK2_PHOSPHO_SITE        7
[PROSITE]       PKC_PHOSPHO_SITE        5
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY     4.11 %

SEQ     MYQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFYLSHREGY
SEG     ............................................................
PRD     cccccccccccccccceeeccccccccccccccccccccccccccccccceeeeeecccc

SEQ     GTMYMKTRLFQTHCHNDIVNLLLDCGADVNKCSDEGLTALSMCFLLHYPAQSFKPNVAER
SEG     ............................................................
PRD     ccccchhhhhheeecccccchhhhhccccccccccccccchhhhhhhhhccccccccccceee

SEQ     TIPEPQEPPKFPVVPILSSSFMDTNLESLYYEVNVPSQGSYELRPPPAPLLLPRVSGSHE
```

```
SEG     .......................................xxxxxxxxxxxx......
PRD     ecccccccccceeeeeeeccccccccccceeeeeeccccccccccccccccccccccccc

SEQ     GGHFQDTGQCGGSIDHRSSSLKGDSPLVKGSLGHVESGLEDVLGDTDRGSLCSAETKFES
SEG     ............................................................
PRD     cccccccccccccccccccccccccceeecccccccccccccccccccccccceeeeecccc

SEQ     NLCVCDFSIELSQAMLERSAQSHSLLKMASPSPCTSSFDKGTMRRMALSMIE
SEG     ....................................................
PRD     ccccchhhhhhhhhhhhhhhhhhhhcccccccccccccccchhhhhhhccc
```

Prosite for DKFZphfbr2_23n16.1

PS00005	55->58	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	200->203	PKC_PHOSPHO_SITE	PDOC00005
PS00005	226->229	PKC_PHOSPHO_SITE	PDOC00005
PS00005	282->285	PKC_PHOSPHO_SITE	PDOC00005
PS00006	55->59	CK2_PHOSPHO_SITE	PDOC00006
PS00006	121->125	CK2_PHOSPHO_SITE	PDOC00006
PS00006	140->144	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	217->221	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00008	45->51	MYRISTYL	PDOC00008
PS00008	86->92	MYRISTYL	PDOC00008
PS00008	177->183	MYRISTYL	PDOC00008
PS00008	188->194	MYRISTYL	PDOC00008
PS00008	229->235	MYRISTYL	PDOC00008
PS01159	19->44	WW_DOMAIN_1	PDOC50020

(No Pfam data available for DKFZphfbr2_23n16.1)

DKFZphfbr2_23o24
----

group: brain derived

DKFZphfbr2_23o24 encodes a novel 139 amino acid protein with similarity to CAAX-box proteins.

The CAAX box is a prenyl group binding site found in a number of eukaryotic proteins, such as which is found in Ras- and ras-like proteins such as Rho, Rab, Rac, Ral, and Rap, as well as in nuclear lamins A and B, some G protein alpha and gamma subunits and some dnaJ-like proteins. These proteins are posttranslationally modified at this site by the attachment of either a farnesyl or a geranyl-geranyl group to a cysteine residue.

No informative BLAST results; no predictive prosite, pfam or SCOP motife

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to lectins

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 3564 bp
Poly A stretch at pos. 3541, no polyadenylation signal found

```
   1 GAATGGCTCC GCAGATGGCC GGCACTGAGA GCCAGCAAGA AGCGGAGGAG
  51 ATGGGCCTTC AGCAGGGGGT TGCGGGGGGA GCTTTAAACT GAGCCCTGTA
 101 AACATGGCAG AACTGCTCAG TGGGAGACTC TCAGCACAGA CGGTCATGGG
 151 GAAGTGAGTG CAGTTCATTT GTAATCTTGT TGTCGAGTTC TGGGTTTTTT
 201 TTGTTTGTTT CGTAACTTTA AAGGTATGCA CTTTATATAG ATTTATTTAT
 251 TTGCTGGGAC CGTTACTCAG AGTTCCTAGA AATGTACACA GCTTTTTTAC
 301 CAGGGTTACT CCTCAGAATC ACTTGTCACT TCTTTAAATG AATGAATGAA
 351 TGTGCCAGGC CCTATGCCTG GAGGTTGGGA GCTTCATCTA CATCACATTC
 401 TAACAGGTGA CCACTGGGGT AAGCACTGTG TGACTGCAAA GCCAGGGTGT
 451 GTTTCCATCA ACACCCAGAT GACCGTGCCT ATGTGCCCCT GTTGTCCTCC
 501 CTCCAGGACT GCCTCCTCAC CCCACCCCTT TCTGCAGCTC CTCATCTAAA
 551 CATCTCGCCT GGTGAGGTCA CGGCTTAGCC TGTTGGCCAG TGGCCCCACC
 601 ACCATCCTTC CCCCTGTGCA GATTGGAGGA GGCCAGGTCT CTCCCCTTAG
 651 CTCCTATGTC CCCTTCACCC CCCATGGCAC AGATGAGACA TTCACAGAGT
 701 TTGCAGATGA TGGAAGAGAA GACTCCAGGT TGCCAGGTGT GTCCACTCTC
 751 AGGAACCCCC AGCCCAAGCC TCACTGCTCG TGTTCCCAGC CAACCCCAGC
 801 ACGGGGGATA CGCCGGTGCT GTTTCCCTGC TCAGATACAA CCAGTTACCA
 851 GAAACGACCT CACCCCTCCA ACCACTTTCC AAGGTGCCAG GACAGAGAAG
 901 CCCTTCACTG GCCCACCCAG GGCAGTTGAC AGAGGGATGC CCTCCTTGGA
 951 GGGGAGCCTC ACCTCTACCC ACAGGGCCGC GGCCTTGTCC TGGATTCTCA
1001 CCGGGGCAGT CACGTCAGGA TGGAGAGGTC CCATGTCAGC CAGTTCTTTG
1051 GTGGGGGTCA TGTAGTCTGA AATGACCTGC CGATGGTCCA GGCTGAGCCA
1101 GGGAAGCTGA GCCTGGGTGC CTTTTTGGTG CCTACTCTGA CTTGAGTTGG
1151 ATTCATGCCA CAGACCCACC TTCTTGAGCA ACAACACATA TAGCCACCAA
1201 CACAAGAGCC AGGCACACAC TGAGCAGAGA AAGTCCCTGT CGCCTCACCA
1251 CCCAAAAACT CCAGCTTTGC AGAGACCAAG GTTCTTCTCT ACCTTTGCAG
1301 AAGCCTCTGT GACCAAACCC GGAGCTTGCC CTTCTGAGGC CTCTAGCATT
1351 TCTCCAGGTG TTTTTCAGAG GACTTGGTTT AAATTTGTTC ACCCCAAATG
1401 TGGTCTTTCC CGGATCATGA AAGGATCTGC CGCAAAGGTG AATCTGAGTC
1451 TCCTCAGAGT CATATGAGAC TGAAACTGCT TATAACATTT CCGTGACCTA
1501 ATAAGTCTTC CAAAAATGTA GGGTATTAAG AGTTTAGTGA CATTAAAAAG
1551 TTTAGTCGAA AATATCGTGA TTCAGGTATA TTTAGACATT TGATTCATGC
1601 CAAATTGCCA CTGTTAACAG AAAACACACC CCAAGCACAT TAATGCCTAG
1651 ATATTTCAAA CCCTTTTCTG CCCACACATT CTTAAAAATA ATATACTGAG
1701 AAATCTATAT ACAGGTTTTT TTTTAATTAG CTTGGAAAAG AGCAGTTGTA
1751 TTCTGTTTGA ACAGCTGCTA ATGTCAATTC CTGTGGGAAG AAAGACCAAA
1801 GAACATGGAG TTACACCAAG AATTTTAAAA CAAAGACGCT GTCCCTTTCC
1851 TGAGCACCGT GCAGCCAAGA CTGAGAGATC AGTCTGAGAC CTGTGATTAA
1901 GGAGTGTTTT CTACATAGCG TATAATTATG GAGCCACACA AGTGGGCCAT
1951 TACTCTGTTG AGTGCTTCAT GTTTGAGGTA TTTTCGTGTT CCAACTTACA
2001 TTAAAGTGTT TATAAAACAG GAAAAATCCA CGAGCAGGTA TTGACACTAT
2051 CCATATTAGA TCATCACAAA ATTATATATA TAGCAGAGTC ATAAACAATG
2101 AGAAACGGTC TTCCCACACT TGCTTTAAAT GGCCATGACC TAGTGTTTAG
2151 GGAAAGCAGT AAAATCAGCG AGGAGCTCGT GGGAAAAATG AGACGGGCCC
2201 TGAGGGGGTG ACTCATGGGC CAAGCAGGGC CACACAGGTA CCAGGCCGCC
2251 ACGTCCTCTC CTGCCTCTCA CTCTCTGGAG ACTGGACTTC CTTTACTGCC
2301 TCCTTTCTGA CATTTCCTAG ACATCAGACT TTGCTACTTA GTACACAAAC
2351 GGGGTTCCCT TTTAAATTTG TTCACTCTAG TTAGCATTTG CAGAAGCTGT
2401 GAAAAATTAC AGAGAGATGA TGTGTTGGGT AAGAGATGGT TTAAAAGTCC
```

```
2451 AGCTTGCTGT TTTTCATTAA GTGTCTTGAA AATGAGTAAG TGGCGTTCCT
2501 GGAGGGGAAC AATCATATAA TTCCGCAGGG TGGGTCTAAA CTTGTTTTCT
2551 GATAGTGTTT AGCAGCTCAT GGCTCTGAGG GCACCTGATA ACACAGCAGC
2601 CAGGCGCTGA TGAGAAGTGT GTGCCAGACA GACCCGAGTG TGGCTTGGCT
2651 CTTGCCTTAT GTTCCTTTCT CTGTTCAGAG AAGCGTGAGA TGAGATTTTG
2701 TGATTATATT GCACTCCTTG GGCTGACTTT CCCATGCACA GAATGTTTTA
2751 CACATCCTGA TAGCTGAGCT GAAAATGCAA AGAGAAGGGA AAATGCCTTA
2801 AATTGTTCTG GCTAATTTAG AAGCAGCAGG CCTTGGAAGT CTTTGTCCTG
2851 TGTCCCTGAA CAAATCTTAT GGGAGCTCTG GTACCTATGC CAGAAAATGC
2901 ACATAGGCAC AACACTTTTA CATACACGTT CACACACCCC ACCCTTATGG
2951 AGAACTTTTT TCTAAATAAG AGAAAGAAAA ATTTTAAGAC TTACAAGTTA
3001 TGTTTAGGTA TTTTACATGG TTCAGAAAAC AAGACATGAA GCGGTATAAA
3051 CTGAGAAGTC TTGTTCCCAC AACCCCACGT GCCAGGTACA CATAACCATT
3101 TTTATTCACC TCTAGCTTGT GCTTCCAATG TTTGTTAGGC ATATGTAAAT
3151 AAGTGAATAG ATAAGCATTT CTCCCTCCTT TTGCTGACAT GAGTGGTGGC
3201 ATGTTTTGCC CCTGGCTTTT ATCCCTTGAC CCCATTCCAG TACCTAGAGA
3251 CCTGCTTCAT TTTTTTAGAT GTGTAATACT TCATGTGTGC GTGTGCCTTA
3301 GTGATTAACT CGTGCACTGT GCAGGGACAT CGGGCTGGGA TCAGTTTGTT
3351 CACTGATATA TACAGCGCTG CGGGAGATAC CCTCACATGT GTATCATTTG
3401 GTCCATGTGC AGGTGTGTCT GGAAGATAGA ATTCTAGGCG TAGAATTGAT
3451 AGGTTAAATG TATTTATAGG GAAAAAATCA ATATAAAACT TTGCGTGTAA
3501 TGATATTTGC GTGCTTTTTT TTTTAATTTT TTTACCCAAA TAGTAAAAAA
3551 AAAAAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 656 bp to 1072 bp; peptide length: 139
Category: similarity to known protein

```
  1 MSPSPPMAQM RHSQSLQMME EKTPGCQVCP LSGTPSPSLT ARVPSQPQHG
 51 GYAGAVSLLR YNQLPETTSP LQPLSKVPGQ RSPSLAHPGQ LTEGCPPWRG
101 ASPLPTGPRP CPGFSPGQSR QDGEVPCQPV LWWGSCSLK
```

BLASTP hits

Entry CEEGAP7_1 from database TREMBL:
gene: "EGAP7.1"; Caenorhabditis elegans cosmid EGAP7.
Score = 123, P = 2.3e-07, identities = 35/103, positives = 44/103

Entry MMBPC35_1 from database TREMBL:
Mouse carbohydrate binding protein 35 mRNA, 3' end.
Score = 113, P = 2.2e-06, identities = 40/103, positives = 44/103

Entry A28651 from database PIR:
galactose-specific lectin - mouse >TREMBL:MMMAC2A_1 Mouse mRNA for Mac-2 antigen
Score = 113, P = 2.2e-06, identities = 40/103, positives = 44/103

Alert BLASTP hits for DKFZphfbr2_23o24, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23o24, frame 2

Report for DKFZphfbr2_23o24.2

[LENGTH]	139	
[MW]	14748.91	
[pI]	8.90	
[PROSITE]	PRENYLATION	1

[PROSITE]	MYRISTYL	1
[PROSITE]	CK2_PHOSPHO_SITE	1
[PROSITE]	PROKAR_LIPOPROTEIN	1
[PROSITE]	PKC_PHOSPHO_SITE	1
[KW]	All_Alpha	

```
SEQ   MSPSPPMAQMRHSQSLQMMEEKTPGCQVCPLSGTPSPSLTARVPSQPQHGGYAGAVSLLR
PRD   ccccchhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccchhhhhhhh

SEQ   YNQLPETTSPLQPLSKVPGQRSPSLAHPGQLTEGCPPWRGASPLPTGPRPCPGFSPGQSR
PRD   hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   QDGEVPCQPVLWWGSCSLK
PRD   ccccccccccccccccccc
```

Prosite for DKFZphfbr2_23o24.2

PS00005	40->43	PKC_PHOSPHO_SITE	PDOC00005
PS00006	119->123	CK2_PHOSPHO_SITE	PDOC00006
PS00008	50->56	MYRISTYL	PDOC00008
PS00013	126->137	PROKAR_LIPOPROTEIN	PDOC00013
PS00294	136->140	PRENYLATION	PDOC00266

(No Pfam data available for DKFZphfbr2_23o24.2)

DKFZphfbr2_23o5

group: brain derived

DKFZphfbr2_23o5 encodes a novel 360 amino acid protein with no known similarity

No informative BLAST results; no predictive prosite, pfam or SCOP motife

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

potential start at Bp 24 matchs Kozak consensus ANNatgG

Sequenced by AGOWA

Locus: /map="7q21-q22"

Insert length: 1736 bp
Poly A stretch at pos. 1714, polyadenylation signal at pos. 1680

```
   1 GGGGGAGGAT CAAAGTAGGC AAGATGGCGT CGAGCGGCGG GGAGCCAGGG
  51 AGTTTATTTG ATCACCACGT CCAGAGGGCG GTATGCGACA CACGGGCCAA
 101 ATATCGAGAG GGACGACGGC CTCGTGCTGT GAAGGTATAT ACAATCAATT
 151 TGGAATCTCA GTACTTATTA ATACAAGGAG TTCCTGCTGT GGGAGTCATG
 201 AAGGAATTAG TTGAGCGATT CGCTTTATAT GGTGCAATTG AACAGTACAA
 251 TGCTCTAGAT GAATACCCAG CAGAAGACTT TACTGAAGTT TATCTTATTA
 301 AATTTATGAA CTTACAAAGT GCAAGGACAG CCAAGAGAAA AATGGATGAA
 351 CAGAGTTTCT TCGGTGGATT GCTTCATGTG TGCTATGCTC CAGAATTTGA
 401 AACAGTTGAA GAAACTAGAA AAAAACTACA AATGCGGAAG GCATATGTAG
 451 TAAAAACTAC TGAAAATAAA GACCATTACG TGACAAAGAA GAAATTGGTT
 501 ACAGAGCATA AAGACACAGA GGATTTTAGA CAAGACTTCC ACTCAGAGAT
 551 GTCTGGATTT TGTAAAGCTG CTTTGAACAC TTCTGCAGGG AACTCAAATC
 601 CTTATCTTCC GTATTCCTGT GAATTGCCTT TATGTTATTT CTCCTCAAAA
 651 TGTATGTGTT CATCCGGGGG ACCTGTAGAC AGAGCACCAG ACTCCTCTAA
 701 GGATGGTAGA AACCATCATA AAACAATGGG GCATTATAAC CACAATGACT
 751 CTTTGCGGAA AACACAGATA AACTCTTTGA AAAACTCAGT GGCCTGCCCT
 801 GGTGCACAAA AGGCTATTAC GTCTTCAGAG GCAGTTGACA GATTTATGCC
 851 TAGGACAACA CAACTGCAGG AGCGCAAAAG AAGAAGAGAA GATGATCGTA
 901 AACTTGGAAC TTTTCTTCAA ACAAACCCAA CTGGTAATGA GATTATGATT
 951 GGACCTCTGT TACCAGACAT CTCTAAAGTG GATATGCACG ATGACTCATT
1001 GAATACAACG GCGAATTTAA TTCGGCATAA ACTTAAAGAG GTATTTCATC
1051 TGTGCCAAAG CCTCCAGAGG ACAAGCCAGA AGATGTACAT ACAAGTCATC
1101 CATTAAAACA AAGAAGAAGA ATATAGAGTG CCAGCAGCAA CTTAGTATTT
1151 TCTAAAAAGA ACATTTATTA TTTATTTTTA GCCTGTCATT TTAATTCTTC
1201 AAGAGATTTT ACTGCTGGTA TTTTTTGATG CACTCCTCTT TGTAATTTCA
1251 TTCAAGCCAT TTGTCTAAAG TCATTTCTTT GTTTTTTGGG AGATGGAGTC
1301 TTGCTCTGTT GCCCAGGCTG GAATGCAGTG GCGTGATCTC GGCTCACTGC
1351 AACCTCCACC TCCCGGGTTC AAGCGATTCT CCTGCCTCAG CCTCCTGAGT
1401 ATCTGGGATT ACAGGCGTGC ACCACCATGC CTGGCTAAGT TTTGTGTTTT
1451 TTTTAGTAGA GATGGGTTTT CACCATATTG GTCAGGCTGG TCTCGAACTC
1501 CTGACCTTGT GATACACCTG CCTCAGCCTC CCAAAGGGAT GAGCCACCGC
1551 GCCTGGCCCA TTTCTTCTTT TTTTGACCCA TACTTAATGT TGCAGAAACT
1601 ATTCTTGTCA TAACATTATC TCTCATGTAC AGTAATTATA TGTAAATTAA
1651 TTGAAGCAAA TATGGAAACT TTACAATAGA AATAAAGATA GGCAGCCAGC
1701 GTCTGTTTCC AATTATAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry AC005156 from database EMBL:
Homo sapiens PAC clone DJ1099C19 from 7q21-q22, complete sequence.
Score = 2897, P = 2.4e-154, identities = 583/586
2 exons covering Bp 465-1723

Medline entries

No Medline entry

Peptide information for frame 3

```
ORF from 24 bp to 1103 bp; peptide length: 360
Category: similarity to unknown protein


  1 MASSGGEPGS LFDHHVQRAV CDTRAKYREG RRPRAVKVYT INLESQYLLI
 51 QGVPAVGVMK ELVERFALYG AIEQYNALDE YPAEDFTEVY LIKFMNLQSA
101 RTAKRKMDEQ SFFGGLLHVC YAPEFETVEE TRKKLQMRKA YVVKTTENKD
151 HYVTKKKLVT EHKDTEDFRQ DFHSEMSGFC KAALNTSAGN SNPYLPYSCE
201 LPLCYFSSKC MCSSGGPVDR APDSSKDGRN HHKTMGHYNH NDSLRKTQIN
251 SLKNSVACPG AQKAITSSEA VDRFMPRTTQ LQERKRRRED DRKLGTFLQT
301 NPTGNEIMIG PLLPDISKVD MHDDSLNTTA NLIRHKLKEV FHLCQSLQRT
351 SQKMYIQVIH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23o5, frame 3

TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC F15K20 genomic sequence, complete sequence., N = 2, Score = 114, P = 3.6e-11

```
>TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II
     BAC F15K20 genomic sequence, complete sequence.
          Length = 227

  HSPs:

 Score = 114 (17.1 bits), Expect = 3.6e-11, Sum P(2) = 3.6e-11
 Identities = 21/41 (51%), Positives = 29/41 (70%)

Query:   103 AKRKMDEQSFFGGLLHVCYAPEFETVEETRKKLQMRKAYVV 143
             AKRK+DE SF G  L + YAPE+E V +T+ KL+ R+  V+
Sbjct:    51 AKRKLDESSFLGNRLQISYAPEYENVNDTKDKLESRRKEVL 91

 Score = 107 (16.1 bits), Expect = 2.6e-10, Sum P(2) = 2.6e-10
 Identities = 50/191 (26%), Positives = 83/191 (43%)

Query:   103 AKRKMDEQSFFGGLLHVCYAPEFETVEETRKKLQMRKAYVVKTTENKDHYVTKKKLVTEH 162
             AKRK+DE SF G  L + YAPE+E V +T+ KL+ R+  V+      +    T +  VT+
Sbjct:    51 AKRKLDESSFLGNRLQISYAPEYENVNDTKDKLESRRKEVLARLNPQKEKSTSQ--VTKL 108

Query:   163 KDTEDFRQDFHSEMSGFCKAALNTSAGNSNPYLPYSCELPLCYFSSKCMCSSGGPVDRAP 222
                   + D  S       +    +   GN+ P     S +     YF+S  M  +   V
Sbjct:   109 AGPALTQTDNVSSQRREMEYQFHR--GNA-PVTRVSSDQE--YFASSMNQTVKTV---- 159

Query:   223 DSSKDGRNHHKTMGHYNHNDSLRKTQINSLKNSVACPGAQKAITSSEAVDRFMPRTTQLQ 282
                K  +   + +  +H   + ++  N  +       P +Q     S     R  P ++Q+Q
Sbjct:   160 -REKLNKTREENISSLSHCKQIEESG-NQKRLQ---PSSQTQPEESGNQKRLQP-SSQIQ 213

Query:   283 -ERKRRREDDRK 293
              + KR R D+R+
Sbjct:   214 PDLKRTRVDNRR 225

 Score = 102 (15.3 bits), Expect = 3.6e-11, Sum P(2) = 3.6e-11
 Identities = 22/55 (40%), Positives = 38/55 (69%)

Query:    26 KYREGRRPRAVKVYTINLESQYLLIQGVPAVGVMKELVERFALYGAIEQY--NALDE 80
             +Y++    P AV+VYT+   ES+Y++++ VPA+G    +L+  F   YG +E++     LDE
Sbjct:     3 RYKD-ETP-AVRVYTVCDESRYMIVRNVPALGCGDDLMRLFMTYGEVEEFAKRKLDE 57
```

Pedant information for DKFZphfbr2_23o5, frame 3

Report for DKFZphfbr2_23o5.3

```
[LENGTH]        360
[MW]            41105.85
[pI]            8.89
[HOMOL]         TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC
F15K20 genomic sequence, complete sequence. 5e-12
[PROSITE]       AMIDATION        1
[PROSITE]       MYRISTYL         2
[PROSITE]       CK2_PHOSPHO_SITE        7
```

[PROSITE]	PKC_PHOSPHO_SITE	9
[PROSITE]	ASN_GLYCOSYLATION	3
[KW]	Alpha_Beta	
[KW]	LOW_COMPLEXITY	4.17 %

```
SEQ   MASSGGEPGSLFDHHVQRAVCDTRAKYREGRRPRAVKVYTINLESQYLLIQGVPAVGVMK
SEG   ............................................................
PRD   ccccccccceeeecceeeeehhhhhhhhhccccceeeeeeecccceeeeeecccccchhhh

SEQ   ELVERFALYGAIEQYNALDEYPAEDFTEVYLIKFMNLQSARTAKRKMDEQSFFGGLLHVC
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhcccccceeeeeeehhhhhhhhhhhhhhhhhhcccccccceeee

SEQ   YAPEFETVEETRKKLQMRKAYVVKTTENKDHYVTKKKLVTEHKDTEDFRQDFHSEMSGFC
SEG   ............................................................
PRD   eccchhhhhhhhhhhhhhhhheeeeccccceeeeeeeeeeecccchhhhhhhhhhcccce

SEQ   KAALNTSAGNSNPYLPYSCELPLCYFSSKCMCSSGGPVDRAPDSSKDGRNHHKTMGHYNH
SEG   ............................................................
PRD   eeeecccccccccccccccccccceeeccccccccccccccccccccccccccccccccc

SEQ   NDSLRKTQINSLKNSVACPGAQKAITSSEAVDRFMPRTTQLQERKRRREDDRKLGTFLQT
SEG   .......................................xxxxxxxxxxxxxxx......
PRD   cccceeeecccccccccccccceeeeecceeeeecccccchhhhhhhhhhhccceeeeeec

SEQ   NPTGNEIMIGPLLPDISKVDMHDDSLNTTANLIRHKLKEVFHLCQSLQRTSQKMYIQVIH
SEG   ............................................................
PRD   cccccceeeecccccccccccccccccccchhhhhhhhhhhhhhhhhhhhcchhhhhhccc
```

Prosite for DKFZphfbr2_23o5.3

PS00001	185->189	ASN_GLYCOSYLATION	PDOC00001
PS00001	241->245	ASN_GLYCOSYLATION	PDOC00001
PS00001	327->331	ASN_GLYCOSYLATION	PDOC00001
PS00005	99->102	PKC_PHOSPHO_SITE	PDOC00005
PS00005	102->105	PKC_PHOSPHO_SITE	PDOC00005
PS00005	131->134	PKC_PHOSPHO_SITE	PDOC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00005	207->210	PKC_PHOSPHO_SITE	PDOC00005
PS00005	224->227	PKC_PHOSPHO_SITE	PDOC00005
PS00005	243->246	PKC_PHOSPHO_SITE	PDOC00005
PS00005	251->254	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00006	4->8	CK2_PHOSPHO_SITE	PDOC00006
PS00006	10->14	CK2_PHOSPHO_SITE	PDOC00006
PS00006	127->131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	224->228	CK2_PHOSPHO_SITE	PDOC00006
PS00006	266->270	CK2_PHOSPHO_SITE	PDOC00006
PS00006	303->307	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006
PS00008	5->11	MYRISTYL	PDOC00008
PS00008	260->266	MYRISTYL	PDOC00008
PS00009	29->33	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_23o5.3)

DKFZphfbr2_2a2
----

group: brain derived

DKFZphfbr2_2a2.3 encodes a novel 167 amino acid protein with weak similarity to human 52K autoantigen Ro/SS-A

The novel protein contains a C3HC4 Zinc finger "RING finger" motive.
This domain is probably involved in mediating protein-protein interactions.
Proteins containing a RING-finger are: mammalian V(D)J recombination activating protein (RAG1), mouse rpt-1, human rfp, human 52 Kd Ro/SS-A protein and others.

No informative BLAST results; no predictive prosite, pfam or SCOP motife

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to 52K autoantigen Ro/SS-A - human

complete cDNA, complete cds, few EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1376 bp
Poly A stretch at pos. 1355, polyadenylation signal at pos. 1340

```
   1 GGGGACTCCA AATTAGAAAG GGGACGTCTA GTGGGTTGCC CGGGAGGGGT
  51 GGCGGGAGCG GTCCTGGAAA TAATCTGTCC TCTGTCGCCG GGAACTGGCG
 101 AGGTAGTTCC TTCGCGGTGG AGAGACCTGG AATGGCCAAA TATCAAGGTG
 151 AAGTTCAAAG TTTGAAACTG GATGATGATT CAGTTATAGA AGGAGTAAGC
 201 GACCAAGTAC TTGTGGCAGT TGTGGTCAGT TTCGCTTTGA TTGCTACCCT
 251 GGTATATGCA CTTTTCAGAA ATGTACATCA AAACATTCAC CCAGAAAACC
 301 AGGAGCTAGT AAGGGTACTT CGAGAACAGC TTCAAACAGA ACAGGATGCA
 351 CCTGCTGCCA CTCGACAGCA GTTCTACACT GACATGTACT GTCCCATCTG
 401 CCTGCACCAA GCCTCCTTCC CGGTGGAGAC CAACTGTGGA CATCTTTTTT
 451 GTGGTGCCTG CATTATTGCT TACTGGCGAT ATGGTTCATG GCTTGGGGCA
 501 ATCAGTTGTC CAATCTGTAG ACAAACGGTA ACCTTACTCC TAACAGTATT
 551 TGGTGAAGAT GATCAGTCTC AGGATGTTCT GAGATTGCAT CAGGATATTA
 601 ATGATTATAA CCGGAGATTC TCAGGGCAAC CCTGATCTAT TATGGAGAGA
 651 ATTATGGATC TACCCACTTT ACTGAGGCAT GCATTCAGGG AAATGTTTTC
 701 AGTCGGGGGC CTTTTCTGGA TGTTTCGCAT CAGGATAATA CTTTGTTTAA
 751 TGGGAGCTTT TTTCTATCTT ATATCACCTC TAGATTTTGT ACCTGAAGCC
 801 TTGTTTGGAA TTCTAGGCTT TCTAGATGAT TTCTTTGTCA TCTTTTTATT
 851 GCTTATCTAC ATCTCTATTA TGTATCGAGA AGTGATAACC CAAAGGCTAA
 901 CTAGATGAAA AAGGAAACAA AACTGAGTTT ACTAGGATAT CTGAGCTAAT
 951 GTAGAACATC AAACAGAAGG ACCCATGGCA GTATAAAGCA ATGAAGCAAT
1001 GGAGTATTAT CTCACAAATA TAAAACCACT ATAAGACAAA CATTTGATTA
1051 TCATTTGACA AATACCTAGG TATAACTGGA ATTTTCATGT TTGAAGTTCT
1101 AATATTAAGT TTAGAATTAT AATGATCTAC AGTTGTATCT TGATTCTATG
1151 TTGTCTGGAA AAAATATGGA ATTATATAAA AAGGGATGCT TTTATATATT
1201 TTTCTTTTCC CCAGAATTAC TTAGATTAAT TAGATGTATA GTAAAATATT
1251 GTTAAATGTC AGTTTATCCA TCTTATCCTT CTCAGCAGGT ACCTATATGA
1301 TAATATATAG CTGTGAAACT CATCTAAATA TTTTTGTTCC AATAAAATAT
1351 TATATACTAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 132 bp to 632 bp; peptide length: 167
Category: similarity to known protein
Classification: unset

Prosite motifs: ZINC_FINGER_C3HC4 (102-112)

```
  1 MAKYQGEVQS LKLDDDSVIE GVSDQVLVAV VVSFALIATL VYALFRNVHQ
 51 NIHPENQELV RVLREQLQTE QDAPAATRQQ FYTDMYCPIC LHQASFPVET
101 NCGHLFCGAC IIAYWRYGSW LGAISCPICR QTVTLLLTVF GEDDQSQDVL
151 RLHQDINDYN RRFSGQP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2a2, frame 3

TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A, N = 1, Score = 194, P = 2e-15

PIR:T05222 hypothetical protein F17I5.130 - Arabidopsis thaliana, N = 1, Score = 159, P = 1.4e-10

TREMBLNEW:AB025011_1 gene: "TRIF"; product: "Trif-d"; Mus musculus mRNA for Trif-d, complete cds., N = 1, Score = 108, P = 2.6e-06

PIR:A37241 52K autoantigen Ro/SS-A - human, N = 1, Score = 115, P = 5e-05

>TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A
Length = 283

HSPs:

Score = 194 (29.1 bits), Expect = 2.0e-15, P = 2.0e-15
Identities = 52/149 (34%), Positives = 78/149 (52%)

```
Query:   16 DSVIEGVSDQVLVAVVVSFALIATLVYALFRNVHQNIHPENQELVRVLREQLQTEQDAPA 75
            D  +E ++ Q+ +A+ V F ++  +  A      Q      E      R Q+ T++
Sbjct:   41 DPDVE-LATQITMAIAVIF-IVKAIFDAWQSRRRQRAASRMDENAE--RNQIITQRRISE 96

Query:   76 ATRQQFYTDMYCPICLHQASFPVETNCGHLFCGACIIAYWRYGSWLGA-ISCPICRQTVT 134
            A  Q  +     CPICL  ASFPV T+CGH+FC  CII YW+    +     C +CR T
Sbjct:   97 ALHQSSHE---CPICLANASFPVLTDCGHIFCCECIIQYWQQSKAIVTPCDCAMCRSTFY 153

Query:  135 LLLTV----FGEDDQSQDVLRLHQ-DINDYNRRFS 164
            +LL V     G  +++ D ++ +   I+DYNRRFS
Sbjct:  154 MLLPVHWPTMGTSEETDDHIQENNIRIDDYNRRFS 188
```

Pedant information for DKFZphfbr2_2a2, frame 3

Report for DKFZphfbr2_2a2.3

```
[LENGTH]     167
[MW]         18941.65
[pI]         4.91
[HOMOL]      TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A 1e-13
[FUNCAT]     06.10 assembly of protein complexes  [S. cerevisiae, YDR265w] 1e-04
[FUNCAT]     30.19 peroxisomal organization       [S. cerevisiae, YDR265w] 1e-04
[FUNCAT]     99 unclassified proteins   [S. cerevisiae, YLR323c] 2e-04
[BLOCKS]     BL00518 Zinc finger, C3HC4 type, proteins
[PROSITE]    ZINC_FINGER_C3HC4     1
[PFAM]       Zinc finger, C3HC4 type (RING finger)
[KW]         Irregular
[KW]         3D
[KW]         LOW_COMPLEXITY     6.59 %
```

```
SEQ    MAKYQGEVQSLKLDDDSVIEGVSDQVLVAVVVSFALIATLVYALFRNVHQNIHPENQELV
SEG    ....................xxxxxxxxxxx.............................
1rmd-  ............................................................

SEQ    RVLREQLQTEQDAPAATRQQFYTDMYCPICLHQASFPVETNCGHLFCGACIIAYWRYGSW
SEG    ............................................................
1rmd-  ..................HHHHHHBTTTTTEETTTEEEETTTEEEEHHHHH---HHHHH

SEQ    LGAISCPICRQTVTLLLTVFGEDDQSQDVLRLHQDINDYNRRFSGQP
```

```
SEG     .....................................................
1rmd-   HCCB-TTTTT...........................................
```

Prosite for DKFZphfbr2_2a2.3

```
PS00518      102->112    ZINC_FINGER_C3HC4       PDOC00449
```

Pfam for DKFZphfbr2_2a2.3

```
HMM_NAME         Zinc finger, C3HC4 type (RING finger)

HMM                  *CPICFcTFQlDyPWPFdePmMlPCgHsFCypCIrrW...........CP
                      CPIC      L+         P++++CGH+FC +CI+ +               CP
Query            87   CPIC-----LHQ---ASFPVETNCGHLFCGACIIAYWRYGSWLGAISCP    127

HMM                  mC*
                     +C
Query           128  IC     129
```

DKFZphfbr2_2b17
----

group: transmembrane protein

DKFZphfbr2_2b17 encodes a novel 285 amino acid protein with similarity to D. melanogaster 30K protein.

The protein contains 3 transmembrane regions.
No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to Drosophila hypothetical 30K protein

complete cDNA, complete cds, EST hits
TRANSMEMBRANE 3

Sequenced by Qiagen

Locus: unknown

Insert length: 1426 bp
Poly A stretch at pos. 1345, polyadenylation signal at pos. 1330

```
   1 GGGGGTATTT CCAAGGACTC CAAAGCGAGG CCGGGGACTG AAGGTGTGGG
  51 TGTCGAGCCC TCTGGCAGAG GGTTAACCTG GGTCAAATGC ACGGATTCTC
 101 ACCTCGTACA GTTACGCTCT CCCGCGGCAC GTCCGCGAGG ACTTGAAGTC
 151 CTGAGCGCTC AAGTTTGTCC GTAGGTCGAG AGAAGGCCAT GGAGGTGCCG
 201 CCACCGGCAC CGCGGAGCTT TCTCTGTAGA GCATTGTGCC TATTTCCCCG
 251 AGTCTTTGCT GCCGAAGCTG TGACTGCCGA TTCGGAAGTC CTTGAGGAGC
 301 GTCAGAAGCG GCTTCCCTAC GTCCCAGAGC CCTATTACCC GGAATCTGGA
 351 TGGGACCGCC TCCGGGAGCT GTTTGGCAAA GATGAACAGC AGAGAATTTC
 401 AAAGGACCTT GCTAATATCT GTAAGACGGC GGCTACAGCA GGCATCATTG
 451 GCTGGGTGTA TGGGGGAATA CCAGCTTTTA TTCATGCTAA ACAACAATAC
 501 ATTGAGCAGA GCCAGGCAGA AATTTATCAT AACCGGTTTG ATGCTGTGCA
 551 ATCTGCACAT CGTGCTGCCA CACGAGGCTT CATTCGTTAT GGCTGGCGCT
 601 GGGGTTGGAG AACTGCAGTG TTTGTGACTA TATTCAACAC AGTGAACACT
 651 AGTCTGAATG TATACCGAAA TAAAGATGCC TTAAGCCATT TTGTAATTGC
 701 AGGAGCTGTC ACGGGAAGTC TTTTTAGGAT AAACGTAGGC CTGCGTGGCC
 751 TGGTGGCTGG TGGCATAATT GGAGCCTTGC TGGGCACTCC TGTAGGAGGC
 801 CTGCTGATGG CATTTCAGAA GTACTCTGGT GAGACTGTTC AGGAAAGAAA
 851 ACAGAAGGAT CGAAAGGCAC TCCATGAGCT AAAACTGGAA GAGTGGAAAG
 901 GCAGACTACA AGTTACTGAG CACCTCCCTG AGAAAATTGA AAGTAGTTTA
 951 CAGGAAGATG AACCTGAGAA TGATGCTAAG AAAATTGAAG CACTGCTAAA
1001 CCTTCCTAGA AACCCTTCAG TAATAGATAA ACAAGACAAG GACTGAAAGT
1051 GCTCTGAACT TGAAACTCAC TGGAGAGCTG AAGGGAGCTG CCATGTCCGA
1101 TGAATGCCAA CAGACAGGCC ACTCTTTGGT CAGCCTGCTG ACAAATTTAA
1151 GTGCTGGTAC CTGTGGTGGC AGTGGCTTGC TCTTGTCTTT TTCTTTTCTT
1201 TTTAACTAAG AATGGGGCTG TTGTACTCTC ACTTTACTTA TCCTTAAATT
1251 TAAATACATA CTTATGTTTG TATTAATCTA TCAATATATG CATACATGAA
1301 TATATCCACC CACCTAGATT TTAAGCAGTA AATAAAACAT TTCGCAAAAG
1351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1401 AAAAAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HSG19630 from database EMBL:
human STS A001T27.
Score = 961, P = 1.2e-36, identities = 193/194

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 189 bp to 1043 bp; peptide length: 285
Category: similarity to unknown protein

```
  1 MEVPPPAPRS FLCRALCLFP RVFAAEAVTA DSEVLEERQK RLPYVPEPYY
 51 PESGWDRLRE LFGKDEQQRI SKDLANICKT AATAGIIGWV YGGIPAFIHA
101 KQQYIEQSQA EIYHNRFDAV QSAHRAATRG FIRYGWRWGW RTAVFVTIFN
151 TVNTSLNVYR NKDALSHFVI AGAVTGSLFR INVGLRGLVA GGIIGALLGT
201 PVGGLLMAFQ KYSGETVQER KQKDRKALHE LKLEEWKGRL QVTEHLPEKI
251 ESSLQEDEPE NDAKKIEALL NLPRNPSVID KQDKD
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2b17, frame 3

PIR:JQ1024 hypothetical 30K protein (DmRP140 5' region) - fruit fly (Drosophila melanogaster), N = 1, Score = 312, P = 6.1e-28

>PIR:JQ1024 hypothetical 30K protein (DmRP140 5' region) - fruit fly (Drosophila melanogaster)
Length = 261

HSPs:

Score = 312 (46.8 bits), Expect = 6.1e-28, P = 6.1e-28
Identities = 68/231 (29%), Positives = 125/231 (54%)

```
Query:    30 ADSEVLEERQKRLPYVPEPYYPESGWDRLRELFGKDEQQRISKDLANICKTAATAGIIGW 89
             AD  V +E +    ++      E+G +RL+++F  DE    I  +L ++ +       +IG
Sbjct:    23 ADEIVDKENKTYKAFLASKPPEETGLERLKQMFTIDEFGSIFSELNSVYQAGFLGFLIGA 82

Query:    90 VYGGIPAFIHAKQQYIEQSQAEIYHNRFDAVQSAHRAATRGFIRYGWRWGWRTAVFVTIF 149
             +YGG+      A   ++E +QA  + + FDA +         T  F + G++WGWR  +F T +
Sbjct:    83 IYGGVTQSRVAYMNFMENNQATAFKSHFDAKKKLQDQFTVNFAKGGFKWGWRVGLFTTSY 142

Query:   150 NTVNTSLNVYRNKDALSHFVIAGAVTGSLFRINVGLRGLVAGGIIGALLGTPVGGLLMAF 209
               + T ++VYR K ++ . ++ AG++TGSL+++++GLRG+ AGGIIG  LG   G    +
Sbjct:   143 FGIITCMSVYRGKSSIYEYLAAGSITGSLYKVSLGLRGMAAGGIIGGFLGGVAGVTSLLL 202

Query:   210 QKYSGETVQERKQKDRKALHELKLEEWKGRLQVTEHLPEKIESSLQEDEPE 260
              K SG +++E            ++  ++K RL   E++ +  +   +++ PE
Sbjct:   203 MKASGTSMEE-----------VRYWQYKWRLDRDENIQQAFKKLTEDENPE 242
```

Pedant information for DKFZphfbr2_2b17, frame 3

Report for DKFZphfbr2_2b17.3

[LENGTH]	285
[MW]	32177.88
[pI]	8.65
[HOMOL]	PIR:JQ1024 hypothetical 30K protein (DmRP140 5' region) - fruit fly (Drosophila melanogaster) 7e-20
[PROSITE]	MYRISTYL 7
[PROSITE]	CK2_PHOSPHO_SITE 5
[PROSITE]	ASN_GLYCOSYLATION 1
[KW]	SIGNAL_PEPTIDE 25
[KW]	TRANSMEMBRANE 3
[KW]	LOW_COMPLEXITY 5.96 %

```
SEQ  MEVPPPAPRSFLCRALCLFPRVFAAEAVTADSEVLEERQKRLPYVPEPYYPESGWDRLRE
SEG  ............................................................
PRD  cccccccceeeeeeeeeehhhhhhhhhhhhhhhhhhhhhhhhhcccccccccchhhhhh
MEM  ............................................................

SEQ  LFGKDEQQRISKDLANICKTAATAGIIGWVYGGIPAFIHAKQQYIEQSQAEIYHNRFDAV
SEG  ............................................................
PRD  hhccchhhhhhhhhhhhhhhhhccceeeecccccchhhhhhhhhhhhhhhhhhhhhhhhh
MEM  ...........MMMMMMMMMMMMMMMMMMMMMMMMMMMMM...................

SEQ  QSAHRAATRGFIRYGWRWGWRTAVFVTIFNTVNTSLNVYRNKDALSHFVIAGAVTGSLFR
SEG  ............................................................
PRD  hhhhhhhhhhhcccccccccceeeeeeeecccccccceeecccccccceeeeecccccceee
MEM  ................MMMMMMMMMMMMMMMMMMMMMMMMMMM................M

SEQ  INVGLRGLVAGGIIGALLGTPVGGLLMAFQKYSGETVQERKQKDRKALHELKLEEWKGRL
```

```
SEG    ..xxxxxxxxxxxxxxxxx.........................................
PRD    eeccccccccccceeeeecccccccchhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhhh
MEM    MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...............................

SEQ    QVTEHLPEKIESSLQEDEPENDAKKIEALLNLPRNPSVIDKQDKD
SEG    .............................................
PRD    cccccccccchhhhhcccccccchhhhhhhhhhcccccceeeccccc
MEM    .............................................
```

Prosite for DKFZphfbr2_2b17.3

PS00001	153->157	ASN_GLYCOSYLATION	PDOC00001
PS00006	53->57	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	253->257	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00008	92->98	MYRISTYL	PDOC00008
PS00008	172->178	MYRISTYL	PDOC00008
PS00008	187->193	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	199->205	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_2b17.3)

DKFZphfbr2_2b5
----

group: cell structure and motility

DKFZphfbr2_2b5 encodes a novel 957 amino acid protein with strong similarity to collagens.

The novel protein contains the typical (xxG)n repeat of collagen proteins and a Pfam von Willebrand factor type A domain. Therefore, the protein seems to be a new collagen alpha chain.

The new protein can find application in modulation of connective tissue, bone and cartilage development and maintainance.

similarity to collagen proteins

shows typical (xxG)n repeat of collagen proteins
[PFAM] von Willebrand factor type A domain

Sequenced by Qiagen

Locus: /map="6"

Insert length: 4160 bp
Poly A stretch at pos. 4141, polyadenylation signal at pos. 4119

```
   1 GGGGGCCCGC TGCAGGGAGA ACGGACTCCG GGCGGAGGGC AGCCAATCCG
  51 TTTCAGCGCA GGTCTTGCTC GGGTTGGGCT TGCCACTGCC TGGAACATAC
 101 CTGTCCCCCT GGCGCAACAC TCAGCTGGCT GCGACCGCAA CCCCGAGCCT
 151 GGACACTGCG CCAGGAATCC TAAAACCAAA ATATTAGAAC GAAAACAGAA
 201 ACATGGCTCA CTATATTACA TTTCTCTGCA TGGTTTTGGT GCTGCTTCTT
 251 CAGAATTCTG TGTTAGCTGA AGATGGGGAA GTAAGATCAA GTTGTCGTAC
 301 TGCTCCGACA GATTTAGTTT TCATCTTAGA TGGCTCTTAT AGTGTTGGCC
 351 CAGAAAACTT TGAAATAGTG AAAAAGTGGC TTGTCAATAT CACAAAAAAC
 401 TTTGACATAG GGCCGAAGTT TATTCAAGTT GGAGTGGTTC AATATAGTGA
 451 CTACCCTGTG CTGGAGATTC CTCTCGGAAG CTATGATTCA GGAGAACATT
 501 TGACGGCAGC AGTGGAATCC ATACTCTACT TAGGAGGAAA CACAAAGACA
 551 GGGAAGGCCA TCCAGTTTGC GCTCGATTAC CTTTTTGACA AGTCCTCACG
 601 ATTTCTGACT AAGATAGCAG TGGTACTTAC GGATGGCAAG TCCCAAGATG
 651 ACGTCAAGGA TGCAGCTCAA GCAGCAAGAG ATAGTAAGAT AACATTATTT
 701 GCTATTGGTG TTGGTTCAGA AACAGAAGAT GCCGAACTTA GAGCTATTGC
 751 CAACAAGCCT TCGTCTACTT ATGTGTTTTA TGTGGAAGAC TATATTGCAA
 801 TATCCAAAAT AAGGGAAGTG ATGAAGCAGA AACTTTGTGA AGAATCTGTC
 851 TGTCCAACAC GAATTCCAGT GGCAGCTCGT GATGAAAGGG GATTTGATAT
 901 TCTTTTGGGT TTAGATGTAA ATAAAAAGGT TAAGAAAAGA ATACAGCTTT
 951 CACCAAAAAA GATAAAAGGA TATGAAGTAA CATCAAAAGT TGATTTATCA
1001 GAACTCACAA GCAATGTTTT CCCAGAAGGT CTTCCTCCAT CATATGTATT
1051 TGTGTCTACT CAAAGATTTA AAGTCAAGAA AATTTGGGAT TTATGGAGAA
1101 TATTAACTAT TGATGGAAGG CCACAAATAG CAGTTACCTT AAATGGTGTG
1151 GACAAAATCT TATTATTTAC AACAACCAGC GTAATTAATG GCTCACAAGT
1201 GGTTACCTTT GCTAACCCTC AAGTTAAGAC GTTGTTTGAT GAAGGCTGGC
1251 ACCAAATTCG TCTCTTAGTA ACAGAACAAG ATGTGACTTT GTATATTGAT
1301 GACCAACAAA TTGAAAACAA GCCCTTACAT CCAGTTTTAG GGATCTTGAT
1351 CAATGGGCAA ACCCAAATTG GAAAATATTC TGGAAAAGAA GAAACTGTTC
1401 AGTTTGATGT CCAAAAGTTG CGAATCTACT GTGACCCAGA ACAGAACAAC
1451 CGGGAGACAG CATGTGAGAT TCCTGGATTT AATGGAGAGT GCCTTAATGG
1501 TCCCAGTGAT GTAGGTTCAA CTCCAGCTCC CTGTATTTGT CCTCCGGGAA
1551 AACCAGGACT TCAAGGCCCC AAAGGTGACC CTGGACTGCC TGGGAACCCT
1601 GGCTACCCTG GACAACCTGG TCAAGATGGT AAGCCTGGAT ATCAGGGAAT
1651 TGCAGGGACA CCAGGTGTTC CAGGATCTCC AGGAATACAA GGAGCTCGAG
1701 GACTACCAGG TTACAAAGGA GAACCAGGGC GAGATGGTGA CAAGGGTGAT
1751 CGTGGACTTC CTGGTTTTCC TGGGCTTCAT GGCATGCCAG GATCAAAGGG
1801 TGAAATGGGT GCCAAAGGAG ACAAAGGATC ACCTGGATTT TATGGCAAAA
1851 AGGGTGCAAA AGGTGAAAAG GGGAATGCTG GCTTCCCTGG CCTCCCTGGA
1901 CCTGCTGGAG AACCAGGAAG ACATGGAAAG GATGGATTAA TGGGTAGTCC
1951 CGGTTTCAAG GGAGAAGCAG GATCCCCTGG TGCTCCGGGG CAGGATGGAA
2001 CACGGGGAGA GCCTGGAATC CCAGGATTTC CTGGAAACCG AGGATTAATG
2051 GGCCAAAAGG GAGAAATTGG GCCTCCAGGA CAGCAAGGAA AAAAAGGAGC
2101 CCCAGGGATG CCTGGTTTAA TGGGAAGCAA TGGCTCACCA GGCCAGCCTG
2151 GAACACCGGG ATCTAAGGGA AGCAAAGGTG AACCTGGAAT TCAAGGGATG
2201 CCTGGGGCTT CAGGGCTCAA GGGAGAACCA GGAGCAACGG GTTCCCCAGG
2251 AGAACCAGGA TACATGGGTT TACCCGGGAT TCAAGGAAAA AAGGGGGACA
2301 AAGGAAATCA AGGTGAAAAA GGTATTCAGG GTCAAAAGGG AGAAAATGGA
2351 AGACAGGGAA TTCCAGGGCA ACAGGGAATT CAAGGCCATC ATGGTGCAAA
2401 AGGAGAGAGA GGTGAAAAGG GAGAACCTGG TGTCCGAGGT GCCATTGGAT
2451 CAAAAGGAGA ATCTGGGGTG GATGGCTTGA TGGGGCCCGC AGGTCCTAAG
2501 GGCAACCTGG GGATCCAGG TCCTCAGGGA CCCCCAGGTT TGGATGGGAA
2551 GCCCGGAAGA GAGTTTTCAG AACAATTTAT TCGACAAGTT TGCACAGATG
```

```
2601 TAATAAGAGC CCAGCTACCA GTCTTACTTC AGAGTGGAAG AATTAGAAAT
2651 TGTGATCATT GCCTGTCCCA ACATGGCTCC CCGGGTATTC CTGGGCCACC
2701 TGGTCCGATA GGCCCAGAGG GTCCCAGAGG ATTACCTGGT TTGCCAGGAA
2751 GAGATGGTGT TCCTGGATTA GTGGGTGTCC CTGGACGTCC AGGTGTCAGA
2801 GGATTAAAAG GCCTACCAGG AAGAAATGGG GAAAAAGGGA GCCAAGGGTT
2851 TGGGTATCCT GGAGAACAAG GTCCTCCTGG TCCCCCAGGT CCAGAGGGCC
2901 CTCCTGGAAT AAGCAAAGAA GGTCCTCCAG GAGACCCAGG TCTCCCTGGC
2951 AAAGATGGAG ACCATGGAAA ACCTGGAATC CAAGGGCAAC CAGGCCCCCC
3001 AGGCATCTGC GACCCATCAC TATGTTTTAG TGTAATTGCC AGAAGAGATC
3051 CGTTCAGAAA AGGACCAAAC TATTAGTGTC TGATGCCTCA TTCAGCAGCC
3101 TAGGCATGGT GCTTTTTCTG TGGTCTTTTG CATCTCAGGA AGATAACCAA
3151 CAGTATCCCT TGAAAAGAAA CTTAAGTACC TCGGTGTTTT TATTTTTTTT
3201 TTCTTATGGA AAAAAATATA AAAGATCACA TATACTGATT TTAAAGGCTC
3251 CTCAGTCATT TGGAGCCCTT GGATTAGCAG CATTAATTAA ATCTCAAGGG
3301 TTTCTTGTAA AGTCCATTTA TGTTAATCAA AGTTGAATAT AAAAATCCAC
3351 CATTGCCTGT TAGCCAGTCA GTTTTAGTCA CTGTGAAATA TTTCACATTC
3401 AGCCTCCATG CAGTAGAGAT TTGAGTTTAA TTTCATGTCC ATGTGACTTT
3451 CATGTTTCCT ATCTCATAGC TCATGCTACT ACATAAGCCA AAACATGTAT
3501 CTCATCATTG GAAGTAAGAT CAGGGCTGAT ATTCACCTGG GATAGACAGT
3551 ATTGGTGAAC TACTCATTTA CTACAGTGTC TCAGCCTTGA TAAAGGGCAG
3601 TGGATTGCCT GTTGTTCGGT GTTGTGAATA GCACCTCTGA ATAAGATTAG
3651 AGTGTTTCTT AATTCATTTC AAACTCTAAA ATTAGATTAA TGGTGGTGCT
3701 AAGAAAGAGT ATTAATTACT TTGGGAATGG TCAAAATTAA CATTAAAAAC
3751 ATTTTAGACA AAAAGTTTCA TTGTACATTC AAAGAAAATG TAAGTTTGGA
3801 AGTACTAAAA GACTATTTTA TACTTGTTGA TTAATCGGAA TGTTTGTTGT
3851 ATGCCTTCAT TTTCCATTTC ACTTATATGT GCATGTCCAT ATATGTTAAT
3901 TTTCATTGTA GCAAAGCTAA TGGAAATAAA GCTAATGCTC TAGTTGAAAG
3951 AAAAGGAAAA CTCCTGAAAT CCTAGAATGT CTTGTTATTT TTAGCTGACT
4001 GTAAAATATT ATGAACAGTC TTTGTGTATT GTGCTTAATG CTTTTGTAAG
4051 AAACAGAATT TGAAATATTT CATCCTTGTC ATGCTCAAAA TTTTGTTACA
4101 TGCTTGTTAT TCAGAGTATA ATAAAGTTTT GTACAGGCCT GAAAAAAAAA
4151 AAAAAAAAAA
```

BLAST Results

Entry HS682J15 from database EMBLNEW:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 682J15
Score = 6240, P = 0.0e+00, identities = 1256/1263
13 exons matching Bp 2015-4118

Entry HS708F5 from database EMBLNEW:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 708F5
Score = 2775, P = 1.0e-221, identities = 739/912
10 exons matching Bp 5-1745

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 203 bp to 3073 bp; peptide length: 957
Category: similarity to known protein

```
  1 MAHYITFLCM VLVLLLQNSV LAEDGEVRSS CRTAPTDLVF ILDGSYSVGP
 51 ENFEIVKKWL VNITKNFDIG PKFIQVGVVQ YSDYPVLEIP LGSYDSGEHL
101 TAAVESILYL GGNTKTGKAI QFALDYLFDK SSRFLTKIAV VLTDGKSQDD
151 VKDAAQAARD SKITLFAIGV GSETEDAELR AIANKPSSTY VFYVEDYIAI
201 SKIREVMKQK LCEESVCPTR IPVAARDERG FDILLGLDVN KKVKKRIQLS
251 PKKIKGYEVT SKVDLSELTS NVFPEGLPPS YVFVSTQRFK VKKIWDLWRI
301 LTIDGRPQIA VTLNGVDKIL LFTTTSVING SQVVTFANPQ VKTLFDEGWH
351 QIRLLVTEQD VTLYIDDQQI ENKPLHPVLG ILINGQTQIG KYSGKEETVQ
401 FDVQKLRIYC DPEQNNRETA CEIPGFNGEC LNGPSDVGST PAPCICPPGK
451 PGLQGPKGDP GLPGNPGYPG QPGQDGKPGY QGIAGTPGVP GSPGIQGARG
501 LPGYKGEPGR DGDKGDRGLP GFPGLHGMPG SKGEMGAKGD KGSPGFYGKK
551 GAKGEKGNAG FPGLPGPAGE PGRHGKDGLM GSPGFKGEAG SPGAPGQDGT
601 RGEPGIPGFP GNRGLMGQKG EIGPPGQQGK KGAPGMPGLM GSNGSPGQPG
651 TPGSKGSKGE PGIQGMPGAS GLKGEPGATG SPGEPGYMGL PGIQGKKGDK
701 GNQGEKGIQG QKGENGRQGI PGQQGIQGHH GAKGERGEKG EPGVRGAIGS
751 KGESGVDGLM GPAGPKGQPG DPGPQGPPGL DGKPGREFSE QFIRQVCTDV
801 IRAQLPVLLQ SGRIRNCDHC LSQHGSPGIP GPPGPIGPEG PRGLPGLPGR
```

```
851 DGVPGLVGVP GRPGVRGLKG LPGRNGEKGS QGFGYPGEQG PPGPPGPEGP
901 PGISKEGPPG DPGLPGKDGD HGKPGIQGQP GPPGICDPSL CFSVIARRDP
951 FRKGPNY
```

BLASTP hits

Entry HSCOL7A1X_1 from database TREMBL:
gene: "COL7A1"; product: "collagen type VII"; Homo sapiens (clones: CW52-2, CW27-6, CW15-2, CW26-5, 11-67) collagen type VII intergenic region and (COL7A1) gene, complete cds.
Score = 949, P = 3.4e-122, identities = 237/553, positives = 281/553

Entry CA17_HUMAN from database SWISSPROT:
COLLAGEN ALPHA 1(VII) CHAIN PRECURSOR (LONG-CHAIN COLLAGEN) (LC COLLAGEN). >TREMBL:HSCOL7A1_1 gene: "COL7A1"; product: "alpha-1 type VII collagen"; Human alpha-1 type VII collagen (COL7A1) mRNA, complete cds.
Score = 949, P = 3.6e-122, identities = 237/553, positives = 281/553

Alert BLASTP hits for DKFZphfbr2_2b5, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2b5, frame 2

Report for DKFZphfbr2_2b5.2

[LENGTH]	957
[MW]	99413.38
[pI]	8.49
[HOMOL]	PIR:A40020 collagen alpha 1(XII) chain precursor - chicken 9e-90
[BLOCKS]	BL01119B Copper-fist domain proteins
[BLOCKS]	BL00313B
[BLOCKS]	BL01113A C1q domain proteins
[BLOCKS]	BL00420A Speract receptor repeat proteins domain proteins
[SCOP]	d1zoob_ 3.45.1.1.1 Integrin CD11a/CD18 (LFA-1) [Human (Hom 2e-58
[SCOP]	d1ido__ 3.45.1.1.2 Integrin CR3 (CD11b/CD18), alpha subunit [Huma 8e-62
[EC]	3.1.1.7 Acetylcholinesterase 7e-24
[PIRKW]	blocked amino end 1e-43
[PIRKW]	duplication 7e-46
[PIRKW]	cornea 1e-35
[PIRKW]	lung 2e-40
[PIRKW]	leukocyte 1e-42
[PIRKW]	skin 1e-40
[PIRKW]	transmembrane protein 1e-37
[PIRKW]	cartilage 3e-59
[PIRKW]	hydroxylysine 4e-62
[PIRKW]	connective tissue 3e-43
[PIRKW]	triple helix 5e-82
[PIRKW]	homotrimer 2e-37
[PIRKW]	bone 6e-40
[PIRKW]	Alport syndrome 1e-42
[PIRKW]	laminin binding 2e-40
[PIRKW]	liver 2e-40
[PIRKW]	glycoprotein 5e-82
[PIRKW]	carboxylic ester hydrolase 7e-24
[PIRKW]	disulfide bond 7e-46
[PIRKW]	cell binding 7e-46
[PIRKW]	heterotrimer 4e-62
[PIRKW]	calcium binding 8e-28
[PIRKW]	alternative splicing 5e-82
[PIRKW]	coiled coil 5e-82
[PIRKW]	basement membrane 7e-46
[PIRKW]	trimer 5e-82
[PIRKW]	pyroglutamic acid 3e-43
[PIRKW]	hydroxyproline 4e-62
[PIRKW]	extracellular matrix 5e-82
[PIRKW]	chondroitin sulfate proteoglycan 6e-41
[PIRKW]	sulfoprotein 7e-39
[PIRKW]	kidney 1e-42
[PIRKW]	angiogenesis inhibitor 6e-36
[PIRKW]	Ehlers-Danlos syndrome 2e-40
[SUPFAM]	fibronectin type III repeat homology 5e-82
[SUPFAM]	scavenger receptor cysteine-rich domain homology 1e-37
[SUPFAM]	C-type lectin homology 6e-30
[SUPFAM]	collagen alpha 2(I) chain 5e-40
[SUPFAM]	collagen alpha 1(I) chain 6e-44

[SUPFAM] fibrillar collagen carboxyl-terminal homology 6e-44
[SUPFAM] animal Kunitz-type proteinase inhibitor homology 2e-38
[SUPFAM] fibronectin type II repeat homology 6e-21
[SUPFAM] complement C1q carboxyl-terminal homology 1e-38
[SUPFAM] collagen alpha 3(VI) chain 2e-31
[SUPFAM] collagen alpha 1(IV) chain 7e-46
[SUPFAM] collagen alpha 1(VI) chain 2e-37
[SUPFAM] von Willebrand factor type C repeat homology 6e-44
[SUPFAM] unassigned collagens 4e-62
[SUPFAM] von Willebrand factor type A repeat homology 5e-82
[SUPFAM] collagen alpha 1(XIV) chain 5e-82
[SUPFAM] pulmonary surfactant protein D 6e-30
[SUPFAM] collagen alpha 1(V) chain 7e-39
[SUPFAM] collagen alpha 1(VIII) chain 1e-38
[SUPFAM] EGF homology 1e-35
[PROSITE] AMIDATION 3
[PROSITE] MYRISTYL 14
[PROSITE] CK2_PHOSPHO_SITE 13
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 2
[PFAM] von Willebrand factor type A domain
[KW] Irregular
[KW] 3D
[KW] SIGNAL_PEPTIDE 23
[KW] LOW_COMPLEXITY 24.24 %

```
SEQ    MAHYITFLCMVLVLLLQNSVLAEDGEVRSSCRTAPTDLVFILDGSYSVGPENFEIVKKWL
SEG    ............................................................
latzB  ................................CCCEEEEEEEEECCCCCCHHHHHHHHHH

SEQ    VNITKNFDIGPKFIQVGVVQYSDYPVLEIPLGSYDSGEHLTAAVESILYLGGNTKTGKAI
SEG    ............................................................
latzB  HHHHHHCCBTTTTEEEEEEEEETTTEEEEEETTTTTTTHHHHHHHHHHCCCCCCCHHHHH

SEQ    QFALDYLFDKSSRFLTKIAVVLTDGKSQDDVKDAAQAARDSKITLFAIGVGSETEDAELR
SEG    ............................................................
latzB  HHHHHHHHCCTTTTTEEEEEEEECCCTTTTHHHHHHHHHHHCEEEEEEEECCCCCHHHHH

SEQ    AIANKPSSTYVFYVEDYIAISKIREVMKQKLCEESVCPTRIPVAARDERGFDILLGLDVN
SEG    ............................................................
latzB  HHHGGGGGGGCECCHHHHHHHHHCHHHHHHHH............................

SEQ    KKVKKRIQLSPKKIKGYEVTSKVDLSELTSNVFPEGLPPSYVFVSTQRFKVKKIWDLWRI
SEG    ............................................................
latzB  ............................................................

SEQ    LTIDGRPQIAVTLNGVDKILLFTTTSVINGSQVVTFANPQVKTLFDEGWHQIRLLVTEQD
SEG    ............................................................
latzB  ............................................................

SEQ    VTLYIDDQQIENKPLHPVLGILINGQTQIGKYSGKEETVQFDVQKLRIYCDPEQNNRETA
SEG    ............................................................
latzB  ............................................................

SEQ    CEIPGFNGECLNGPSDVGSTPAPCICPPGKPGLQGPKGDPGLPGNPGYPGQPGQDGKPGY
SEG    ..........................xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
latzB  ............................................................

SEQ    QGIAGTPGVPGSPGIQGARGLPGYKGEPGRDGDKGDRGLPGFPGLHGMPGSKGEMGAKGD
SEG    xx..........................................................
latzB  ............................................................

SEQ    KGSPGFYGKKGAKGEKGNAGFPGLPGPAGEPGRHGKDGLMGSPGFKGEAGSPGAPGQDGT
SEG    .......xxxxxxxxxxxxx........................................
latzB  ............................................................

SEQ    RGEPGIPGFPGNRGLMGQKGEIGPPGQQGKKGAPGMPGLMGSNGSPGQPGTPGSKGSKGE
SEG    ................xxxxxxxxxxxxxxxxxxxxxx......................
latzB  ............................................................

SEQ    PGIQGMPGASGLKGEPGATGSPGEPGYMGLPGIQGKKGDKGNQGEKGIQGQKGENGRQGI
SEG    ........................................xxxxxxxxxxxxxxxxxxxx
latzB  ............................................................

SEQ    PGQQGIQGHHGAKGERGEKGEPGVRGAIGSKGESGVDGLMGPAGPKGQPGDPGPQGPPGL
SEG    xxxxxxxxxxx.............................xxxxxxxxxxxxxxxxxxxx
latzB  ............................................................

SEQ    DGKPGREFSEQFIRQVCTDVIRAQLPVLLQSGRIRNCDHCLSQHGSPGIPGPPGPIGPEG
SEG    xxxxx.......................................xxxxxxxxxxxxxxxx
```

```
latzB   ............................................................

SEQ     PRGLPGLPGRDGVPGLVGVPGRPGVRGLKGLPGRNGEKGSQGFGYPGEQGPPGPPGPEGP
SEG     xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx......xxxxxxxxxxxxxxxxxxxx
latzB   ............................................................

SEQ     PGISKEGPPGDPGLPGKDGDHGKPGIQGQPGPPGICDPSLCFSVIARRDPFRKGPNY
SEG     xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx......................
latzB   .........................................................
```

Prosite for DKFZphfbr2_2b5.2

PS00001	62->66	ASN_GLYCOSYLATION	PDOC00001
PS00001	329->333	ASN_GLYCOSYLATION	PDOC00001
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	131->134	PKC_PHOSPHO_SITE	PDOC00005
PS00005	250->253	PKC_PHOSPHO_SITE	PDOC00005
PS00005	260->263	PKC_PHOSPHO_SITE	PDOC00005
PS00005	286->289	PKC_PHOSPHO_SITE	PDOC00005
PS00005	393->396	PKC_PHOSPHO_SITE	PDOC00005
PS00005	811->814	PKC_PHOSPHO_SITE	PDOC00005
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	261->265	CK2_PHOSPHO_SITE	PDOC00006
PS00006	343->347	CK2_PHOSPHO_SITE	PDOC00006
PS00006	357->361	CK2_PHOSPHO_SITE	PDOC00006
PS00006	393->397	CK2_PHOSPHO_SITE	PDOC00006
PS00006	419->423	CK2_PHOSPHO_SITE	PDOC00006
PS00006	531->535	CK2_PHOSPHO_SITE	PDOC00006
PS00006	600->604	CK2_PHOSPHO_SITE	PDOC00006
PS00006	657->661	CK2_PHOSPHO_SITE	PDOC00006
PS00006	681->685	CK2_PHOSPHO_SITE	PDOC00006
PS00006	750->754	CK2_PHOSPHO_SITE	PDOC00006
PS00006	754->758	CK2_PHOSPHO_SITE	PDOC00006
PS00008	92->98	MYRISTYL	PDOC00008
PS00008	112->118	MYRISTYL	PDOC00008
PS00008	236->242	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	380->386	MYRISTYL	PDOC00008
PS00008	494->500	MYRISTYL	PDOC00008
PS00008	527->533	MYRISTYL	PDOC00008
PS00008	596->602	MYRISTYL	PDOC00008
PS00008	638->644	MYRISTYL	PDOC00008
PS00008	650->656	MYRISTYL	PDOC00008
PS00008	653->659	MYRISTYL	PDOC00008
PS00008	665->671	MYRISTYL	PDOC00008
PS00008	743->749	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00009	547->551	AMIDATION	PDOC00009
PS00009	628->632	AMIDATION	PDOC00009
PS00009	694->698	AMIDATION	PDOC00009

Pfam for DKFZphfbr2_2b5.2

```
HMM_NAME      von Willebrand factor type A domain

HMM               *DIVFLIDGSdSIGpqNFNrMKDFIeRMMERMDIgPDwIRVGVVQYSdNP
                   D+VF++DGS S+GP NF+++K+ ++++  ++DIGP+ I+VGVVQYSD P
Query          37  DLVFILDGSYSVGPENFEIVKKWLVNITKNFDIGPKFIQVGVVQYSDYP    85

HMM               RqEmrFmFNDYQNKeEILQaIqqMMyWMgggTNTGeAIQYVvrNMFweer
                    E  +++  Y + E++++A+  ++ ++GG T+TG AIQ++++++F +++
Query          86 VLE--IPLGSYDSGEHLTAAVESIL-YLGGNTKTGKAIQFALDYLFDKSS    132

HMM               GmRWenvPQVMIIITDGRSQDDIRDpIneMrrmaGIqvFaIGIGNhDNnn
                  +       +++++++TDG+SQDD++D+++++R+    I+ FAIG+G
Query         133 RF----LTKIAVVLTDGKSQDDVKDAAQAARD-SKITLFAIGVGSETE--    175

HMM               WeELReIASePdEdHVFyVdDFeeLdnMqeqL*
                   +ELR IA++P++ +VFYV+D+ +++ ++E +
Query         176 DAELRAIANKPSSTYVFYVEDYIAISKIREVM    207
```

DKFZphfbr2_2c1
----

group: brain derived

DKFZphfbr2_2c1 encodes a novel 697 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 3973 bp
Poly A stretch at pos. 3914, polyadenylation signal at pos. 3900

```
   1 GGGGGGATTT CGGCGGCGGA AACATGGCGG TCGCGGCCGG GCCGGTAACG
  51 GAGAAAGTTT ACGCCGACAC TGGCCTGTAT TAGCGCGTAT GGCCTCGGGC
 101 CCTCGTTCCC CAAGGCGTGC CGCCTCCCTG TTCTCAGTCG CAGGCTGAAG
 151 CCTTGTCTGC TCTCCTCCTT TTTGGTTTGG TTTTGGAACT GACTCCGAGG
 201 GTTGGGAGAG CGCGTTGGTG GCGACGGCCG AGTCAGATCA CTATAAACAA
 251 AATTTCCACA AGAGAAAATG TTGAAATAGG AGTTGCGGAT ACATTGGATA
 301 TACTGGATGA AATACAAGCG GTTAATTTTT GTAACGTGAG GGAAAAGCCC
 351 ACATTGCTGG TTACATGTGT AAATCACTGC GTTATTGCTT TAGTCATTGT
 401 CTCTATTTAG CAATGACAAG ACTGGAAGAA GTAAATAGAG AAGTGAACAT
 451 GCATTCTTCA GTGCGGTATC TTGGCTATTT AGCCAGAATC AATTTATTGG
 501 TTGCTATATG CTTAGGTCTA TACGTAAGAT GGGAAAAAAC AGCAAATTCC
 551 TTAATTTTGG TAATTTTTAT TCTTGGTCTT TTTGTTCTTG GAATCGCCAG
 601 CATACTCTAT TACTATTTTT CAATGGAAGC AGCAAGTTTA AGTCTCTCCA
 651 ATCTTTGGTT TGGATTCTTG CTTGGCCTCC TATGTTTTCT TGATAATTCA
 701 TCCTTTAAAA ATGATGTAAA AGAAGAATCA ACCAAATATT TGCTTCTAAC
 751 ATCCATAGTG TTAAGGATAT TGTGCTCTCT GGTGGAGAGA ATTTCTGGCT
 801 ATGTCCGTCA TCGGCCCACT TTACTAACCA CAGTTGAATT TCTGGAGCTT
 851 GTTGGATTTG CCATTGCCAG CACAACTATG TTGGTGGAGA AGTCTCTGAG
 901 TGTCATTTTG CTTGTTGTAG CTCTGGCTAT GCTGATTATT GATCTGAGAA
 951 TGAAATCTTT CTTAGCTATT CCAAACTTAG TTATTTTTGC AGTTTTGTTA
1001 TTTTTTTCCT CATTGGAAAC TCCCAAAAAT CCGATTGCTT TTGCGTGTTT
1051 TTTTATTTGC CTGATAACTG ATCCTTTCCT TGACATTTAT TTTAGTGGAC
1101 TTTCAGTAAC TGAAAGATGG AAACCCTTTT TGTACCGTGG AAGAATTTGC
1151 AGAAGACTTT CAGTCGTTTT TGCTGGAATG ATTGAGCTTA CATTTTTTAT
1201 TCTTTCCGCA TTCAAACTTA GAGACACTCA CCTCTGGTAT TTTGTAATAC
1251 CTGGCTTTTC CATTTTTGGA ATTTTCAGGA TGATTTGTCA TATTATTTTT
1301 CTTTTAACTC TTTGGGGATT CCATACCAAA TTAAATGACT GCCATAAAGT
1351 ATATTTTACT CACAGGACAG ATTACAATAG CCTTGATAGA ATCATGGCAT
1401 CCAAAGGGAT GCGCCATTTT TGCTTGATTT CAGAGCAGTT GGTGTTCTTT
1451 AGTCTTCTTG CAACAGCGAT TTTGGGAGCA GTTTCCTGGC AGCCAACAAA
1501 TGGAATTTTC TTGAGCATGT TCCTAATCGT TTTGCCATTG GAATCCATGG
1551 CTCATGGGCT CTTCCATGAA TTGGGTAACT GTTTAGGAGG AACATCTGTT
1601 GGATATGCTA TTGTGATTCC CACCAACTTC TGCAGTCCTG ATGGTCAGCC
1651 AACACTGCTT CCCCCAGAAC ATGTACAGGA GTTAAATTTG AGGTCTACTG
1701 GCATGCTCAA TGCTATCCAA AGATTTTTTG CATATCATAT GATTGAGACC
1751 TATGGATGTG ACTATTCCAC AAGTGGACTG TCATTTGATA CTCTGCATTC
1801 CAAACTAAAA GCTTTCCTCG AACTTCGGAC AGTGGATGGA CCCAGACATG
1851 ATACGTATAT TTTGTATTAC AGTGGGCACA CCCATGGTAC AGGAGAGTGG
1901 GCTCTAGCAG GTGGAGATAC ACTACGCCTT GACACACTTA TAGAATGGTG
1951 GAGAGAAAAG AATGGTTCCT TTTGTTCCCG GCTTATTATC GTATTAGACA
2001 GCGAAAATTC AACCCCTTGG GTGAAAGAAG TGAGGAAAAT TAATGACCAG
2051 TATATTGCAG TGCAAGGAGC AGAGTTGATA AAAACAGTAG ATATTGAAGA
2101 AGCTGACCCG CCACAGCTAG GTGACTTTAC AAAAGACTGG GTAGAATATA
2151 ACTGCAACTC CTGTAATAAC ATCTGCTGGA CTGAAAAGGG ACGCACAGTG
2201 AAAGCAGTAT ATGGTGTGTC AAAACGGTGG AGTGACTACA CTCTGCATTT
2251 GCCAACGGGA AGCGATGTGG CCAAGCACTG GATGTTACAC TTTCCTCGTA
2301 TTACATATCC CCTAGTGCAT TTGGCAAATT GGTTATGCGG TCTGAACCTT
2351 TTTTGGATCT GCAAAACTTG TTTTAGGTGC TTGAAAAGAT TAAAAATGAG
2401 TTGGTTTCTT CCTACTGTGC TGGACACAGG ACAAGGCTTC AAACTTGTCA
2451 AATCTTAATT TGGACCCCAA AGCGGGATAT TAATAAGCAC TCATACTACC
2501 AATTATCACT AACTTGCCAT TTTTTGTATG CTGTATTTTT ATTTGTGGAA
2551 AATACCTTGC TACTTCTGTA GCTGCTCTCA CTTTGTCTTT TCTTAAGTAA
2601 TTATGGTATA TATAAGGCGT TGGGAAAAAA CATTTTATAA TGAAAGTATG
2651 TAGGGAGTCA AATGCTTACT GTAAATGCAT AAGAGACGTT AAAAATAACA
2701 CTGCACTTTC AGGAATGTTT GCTTATGGTC CTGATTAGAA AGAAACAGTT
```

```
2751 GTCTATGCTC TGCAATGGTC AATGATGAAT TACTAATGCC TTATTTTCTA
2801 GGCATATAAT AATAGTTTAG AGAATGTAGA CCAGATAAAT TTGTTTACTG
2851 TTTTAAGAAA ACTACCAGTT TACTTACAGA AGATTCTTTT TTCCAAACAG
2901 TAGGTTTCAT CCAAGACCAT TTGAAGAACT GCAAACTCTT TCTCTTAGAA
2951 AAGAAAGAGG GCAGCCTAAA ATAAACGCAA AATTTGCTTA TACTCCATCA
3001 CATTCAGATG TCTTGGTTGT GACTTATTAC CAGTGTGGCA GAGAACCCAA
3051 GTTACATTTT AGATCAAAAT ATTCTTTATG TAGGTATTGT TAAAAGGCTA
3101 GAGCCTACAA GTTGCTCTTC CATGCGTTGG TCAGGGGGCC CTGAAAACAC
3151 TGGTAATATT AAGAGTCTTT CTCAGGGTAA CTTAATGTTT TCTTAATGAA
3201 CAGTGTTTCC AGCTACAAAT TCTTCCAATA AATTGTCTTC CTTTTTGAAA
3251 AGTACTCTCA TAGAAGAAAT TTAGCAATTT CTCGTTGACT GACTCAGTCT
3301 ATTTTAAGTA TTCAGAAAAG ATTTTGATCC CCATTGAGTT AATGCTCTGC
3351 CTTGAAAATT ATTTTTCTGA TCCTTGTTAG TGATAACATT TTTTTTCTAC
3401 TGAAGGTCAG AGGATAGGAA ACAAGTATTT CTCTTCTGGT ATACATGTAA
3451 TGTATTCTGT AAAAAAGTAT TCATATTGGC AATTTTAGTT AGGCATAAATA
3501 TTGTGGTTGT AATTTTTAAA ACTTAGTGTT TTGTCTGATT AAAGCAGGCA
3551 CTGATCAGGG TATCTCCTAA GAGGTAATTC ACTTCTTATT CCTTTCCAAT
3601 AATTATTACA TTCTAAATTT TCATCTATGA GAAATAACAA ACAAGAAGGG
3651 AATAGAATTA AATTGGGGTA TAATCTAATC TTCATTGTTT AAATGGTTTG
3701 CCTTCTCACC ATTGAAGCCA TTTTTTTATA GCCTCAGAAA GAGGAAATAA
3751 TGCCTCCACC ATTTTCTACC TGGTGACTTG AAAATTGAAC TTTTAAGTTA
3801 GGAAGAAGTT AGAGTCAGGG AACTTGTATA CCACTATCTA TGCAGCATTG
3851 TTATAGTCTG ATTATTTCTG TGTTTTGAAT ATGATTTTCC TAATGCTCTA
3901 AATAAAATTT TGTTAAAAAT CAAAAAAAAA AAAAAAAAAA CTTATCGATA
3951 CCGTCGACCT CGATGATGTC GAC
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 365 bp to 2455 bp; peptide length: 697
Category: putative protein
Classification: unset

```
  1 MCKSLRYCFS HCLYLAMTRL EEVNREVNMH SSVRYLGYLA RINLLVAICL
 51 GLYVRWEKTA NSLILVIFIL GLFVLGIASI LYYYFSMEAA SLSLSNLWFG
101 FLLGLLCFLD NSSFKNDVKE ESTKYLLLTS IVLRILCSLV ERISGYVRHR
151 PTLLTTVEFL ELVGFAIAST TMLVEKSLSV ILLVVALAML IIDLRMKSFL
201 AIPNLVIFAV LLFFSSLETP KNPIAFACFF ICLITDPFLD IYFSGLSVTE
251 RWKPFLYRGR ICRRLSVVFA GMIELTFFIL SAFKLRDTHL WYFVIPGFSI
301 FGIFRMICHI IFLLTLWGFH TKLNDCHKVY FTHRTDYNSL DRIMASKGMR
351 HFCLISEQLV FFSLLATAIL GAVSWQPTNG IFLSMFLIVL PLESMAHGLF
401 HELGNCLGGT SVGYAIVIPT NFCSPDGQPT LLPPEHVQEL NLRSTGMLNA
451 IQRFFAYHMI ETYGCDYSTS GLSFDTLHSK LKAFLELRTV DGPRHDTYIL
501 YYSGHTHGTG EWALAGGDTL RLDTLIEWWR EKNGSFCSRL IIVLDSENST
551 PWVKEVRKIN DQYIAVQGAE LIKTVDIEEA DPPQLGDFTK DWVEYNCNSC
601 NNICWTEKGR TVKAVYGVSK RWSDYTLHLP TGSDVAKHWM LHFPRITYPL
651 VHLANWLCGL NLFWICKTCF RCLKRLKMSW FLPTVLDTGQ GFKLVKS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2c1, frame 2

PIR:A71148 hypothetical protein PH0395 - Pyrococcus horikoshii, N = 1,
Score = 96, P = 0.12

>PIR:A71148 hypothetical protein PH0395 - Pyrococcus horikoshii
Length = 288

HSPs:

Score = 96 (14.4 bits), Expect = 1.3e-01, P = 1.2e-01
Identities = 59/234 (25%), Positives = 116/234 (49%)

```
Query:   77 IASILYYYFSMEAASLSLSNLWFGFLL--GL--LCFLDNSSFKNDVKEESTKYLLLTSIV 132
            ++ +LYY F+  A    ++  L  G+LL    +  L +L N +     V+ +  K  +  ++
Sbjct:   57 LSLVLYYLFAFSALK-TIIFLALGYLLMNSIYELGYLMNDTISRRVEGKVHKVRVKLTVF 115

Query:  133 LRILCSLVERISGYVRHRPTLLTTVEFLELVGFAIASTTMLVEKSLSVILLVVALAMLII 192
              +L +L    I  YV       ++ T+ FL+LVG    ++   +L E +L ++ L+  L    +
Sbjct:  116 DSLLIALSRAI--YV-----VIFTLVFLKLVGLQYSTQVILAEVTLFLVFLLYDLTPKHV 168

Query:  193 DLRMKSFLAIPNLVIFAVLLFFSSLET-PKNPIAFACFFICLITDPFLDIYFSGLSVTER 251
               M SF  +  +  F +LL F    T  +N I  +  FI  I   F   ++   +  +
Sbjct:  169 RTVMLSF-PLKFMKAFVLLLPFIITGTLVENVITLS--FILPIAVRFSQAHYLKTACKDN 225

Query: -252 WKPFLYRGRICRRLSVVFAGMIEL-TFFILSAFK-LRDTHLW-YFVIPGFSIFGIFRMIC 308
              P  ++ R+  R S+++  +  L TF +L +F  L +T L   ++IP F++  +  ++
Sbjct:  226 -PPRDFKRRV-ERFSMMYLQVTSLSTFTVLVSFVYLGNTDLLRQYLIP-FAVNVVLILLS 282

Query:  309 HI 310
            ++
Sbjct:  283 YL 284
```

Pedant information for DKFZphfbr2_2c1, frame 2

Report for DKFZphfbr2_2c1.2

```
[LENGTH]        697
[MW]            79741.46
[pI]            8.41
[KW]            TRANSMEMBRANE  11
[KW]            LOW_COMPLEXITY      9.76 %
```

```
SEQ   MCKSLRYCFSHCLYLAMTRLEEVNREVNMHSSVRYLGYLARINLLVAICLGLYVRWEKTA
SEG   ............................................................
PRD   ccceeehhhhhhhhhhhhhhhhhhhhhhccceeeehhhhhhhhhhhhhhhhhhhcccccc
MEM   ...................................MMMMMMMMMMMMMMMMM........

SEQ   NSLILVIFILGLFVLGIASILYYYFSMEAASLSLSNLWFGFLLGLLCFLDNSSFKNDVKE
SEG   ..xxxxxxxxxxxxxxx.............xxxxxxxxxxxxxxxxxxxxxxx.......
PRD   ccceeeecccccchhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccc
MEM   ...MMMMMMMMMMMMMMMMM...........MMMMMMMMMMMMMMMMM............

SEQ   ESTKYLLLTSIVLRILCSLVERISGYVRHRPTLLTTVEFLELVGFAIASTTMLVEKSLSV
SEG   ...............................xxxxxxxxxxxx.............xxxx
PRD   cchhhhhhhhhhhhhhhhhhhhhhhhhceeeecccccchhhhhhhhhhhhhhhhhhhhhhh
MEM   ....MMMMMMMMMMMMMMMMM....................................MMM

SEQ   ILLVVALAMLIIDLRMKSFLAIPNLVIFAVLLFFSSLETPKNPIAFACFFICLITDPFLD
SEG   xxxxxxxxxxxxxx..............................................
PRD   hhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhccccccccccchhhhhhhhhcccccee
MEM   MMMMMMMMMMMMMM...MMMMMMMMMMMMMMMMM........MMMMMMMMMMMMMMMMM.

SEQ   IYFSGLSVTERWKPFLYRGRICRRLSVVFAGMIELTFFILSAFKLRDTHLWYFVIPGFSI
SEG   ............................................................
PRD   eeeccccccccccceeeccccccccchhhhhhhhhhhhhhhhhhcccccceeeeeeeccccc
MEM   ..........................MMMMMMMMMMMMMMMMM................M

SEQ   FGIFRMICHIIFLLTLWGFHTKLNDCHKVYFTHRTDYNSLDRIMASKGMRHFCLISEQLV
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhccccccceeeeeeeccccccchhhhhhhccccchhhhhhhhhhh
MEM   MMMMMMMMMMMMMMMMM.........................................MM

SEQ   FFSLLATAILGAVSWQPTNGIFLSMFLIVLPLESMAHGLFHELGNCLGGTSVGYAIVIPT
SEG   ............................................................
PRD   hhhhhhhhhhhhcccccccchhhhhhheeehhhhhhhhhhccccccccccccceeeeeeec
MEM   MMMMMMMMMMMMMMM....MMMMMMMMMMMMMMMMM........................

SEQ   NFCSPDGQPTLLPPEHVQELNLRSTGMLNAIQRFFAYHMIETYGCDYSTSGLSFDTLHSK
SEG   ............................................................
PRD   cccccccccccccccccccccccccchhhhhhhhhhhhhhhhhccccccccccccchhhhhh
MEM   ............................................................

SEQ   LKAFLELRTVDGPRHDTYILYYSGHTHGTGEWALAGGDTLRLDTLIEWWREKNGSFCSRL
SEG   ............................................................
PRD   hhhhhhhhcccccccceeeeeeeccccccccccceeeccccchhhhhhhhhhhhhcccccceeee
MEM   ............................................................

SEQ   IIVLDSENSTPWVKEVRKINDQYIAVQGAELIKTVDIEEADPPQLGDFTKDWVEYNCNSC
```

```
SEG     ............................................................
PRD     eeeeeccccccchhhhhhccceeeecccedeeeeeeeeccccccccccccccceeeeccccc
MEM     ............................................................

SEQ     NNICWTEKGRTVKAVYGVSKRWSDYTLHLPTGSDVAKHWMLHFPRITYPLVHLANWLCGL
SEG     ............................................................
PRD     cceeeecccceeeeeeeeeccccccceeeecccccchhhhhhhccccccchhhhhhhhhcc
MEM     ............................................................

SEQ     NLFWICKTCFRCLKRLKMSWFLPTVLDTGQGFKLVKS
SEG     .....................................
PRD     eeeeeehhhhhhhhhhhhhcceeeecccccccccccc
MEM     .....................................
```

(No Prosite data available for DKFZphfbr2_2c1.2)

(No Pfam data available for DKFZphfbr2_2c1.2)

DKFZphfbr2_2c17
----

group: signal transduction

DKFZphfbr2_2c17.3 encodes a novel 446 amino acid protein with similarity to yeast YMR131c and mammalian retinoblastoma-binding protein RbAp46

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to YMR131c and retinoblastoma-binding protein RbAp46

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2248 bp
Poly A stretch at pos. 2230, polyadenylation signal at pos. 2200

```
   1 TGGGGAAGAT GGCGGCGCGC AAGGGTCGGC GTCGCACGTG TGAAACCGGG
  51 GAACCCATGG AAGCCGAGTC CGGCGACACA AGTTCCGAGG GCCCGGCCCA
 101 GGTCTACCTG CCCGGCCGGG GGCCGCCGCT ACGCGAAGGG GAGGAGCTGG
 151 TCATGGACGA GGAGGCCTAT GTGCTCTACC ACCGAGCGCA GACTGGCGCC
 201 CCCTGTCTCA GCTTTGACAT AGTCCGGGAT CACCTGGGAG ACAACCGGAC
 251 AGAGCTTCCT CTTACACTTT ACTTGTGTGC TGGGACCCAG GCTGAGAGCG
 301 CCCAGAGCAA CAGACTGATG ATGCTTCGGA TGCACAATCT GCATGGGACA
 351 AAGCCCCCAC CCTCAGAGGG CAGTGATGAA GAAGAAGAGG AGGAAGATGA
 401 AGAGGATGAA GAAGAGCGGA AACCTCAGCT GGAGCTGGCC ATGGTGCCCC
 451 ACTATGGTGG CATCAACCGA GTTCGGGTGT CATGGCTGGG TGAAGAGCCT
 501 GTGGCTGGGG TGTGGTCAGA GAAGGGCCAG GTGGAGGTGT TTGCGCTGCG
 551 GCGGCTTCTG CAGGTGGTGG AGGAGCCCCA GGCCCTGGCA GCCTTCCTCC
 601 GGGATGAGCA GGCCCAAATG AAGCCCATCT TCTCCTTCGC TGGACACATG
 651 GGCGAGGGCT TTGCCCTTGA CTGGTCCCCC CGGGTGACCG GTCGCCTGCT
 701 GACCGGTGAC TGTCAAAAGA ACATCCACCT CTGGACACCT ACGGACGGCG
 751 GCTCCTGGCA CGTGGACCAG CGGCCATTCG TGGGCCACAC ACGCTCTGTG
 801 GAGGACCTGC AGTGGTCACC GACTGAGAAC ACGGTGTTTG CCTCCTGCTC
 851 AGCTGACGCC TCCATCCGCA TCTGGGACAT CCGGGCAGCC CCCAGCAAGG
 901 CCTGCATGCT CACCACAGTC ACCGCCCATG ATGGGGACGT CAATGTCATC
 951 AGCTGGAGCC GCCGGGAGCC CTTCCTGCTC AGTGGCGGGG ATGATGGGGC
1001 CCTCAAGATC TGGGACCTTC GGCAGTTCAA GTCTGGTTCC CCAGTGGCCA
1051 CCTTCAAGCA GCACGTGGCC CCCGTGACCT CCGTCGAGTG GCACCCCCAG
1101 GACAGCGGGG TCTTTGCAGC CTCGGGTGCA GACCACCAGA TCACACAGTG
1151 GGACCTGGCA GTGGAGCGGG ACCCTGAGGC GGGCGACGTG GAGGCCGACC
1201 CCGGACTGGC CGACCTCCCG CAGCAGCTGC TGTTCGTGCA CCAGGGCGAG
1251 ACCGAGCTGA AGGAGCTGCA CTGGCACCCG CAGTGCCCAG GGCTCCTGGT
1301 CAGCACGGCG CTGTCAGGCT TCACCATCTT CCGCACCATC AGCGTCTGAG
1351 GCGTCCCACT GGCTCTGATC TTGCTTCCTG CTTGGAAACT GAAGTCGAAT
1401 TGGGCTCCCC TGGAAGGGGT TCATTCAGGT CTGTTGACTG AGACTGGCCG
1451 GCCTGTGGGC TGCCGTGATG GATTCTGTTT GACGTATTGT TCTCTAGAAG
1501 GCCTGGCTCT GATCCAGTGA CCCCTCTCAC CAAAGAACTC GGTTTAACCA
1551 GGGCTCTGTA AGACCACTCC CACCCAGAGA CTTGTGTGGC CTGGTGTGGC
1601 CTGTGTGTCG GATTCCTTCC TGTCAGCTGT GACCCATTTG ACCTGTGTCC
1651 CCAGAACCCA GTTTTTTGTT TGTTTGTTTG AGACGGAGTC TTGGTCTGTC
1701 GCCCAGGCTG GAGTGCAGTA GCACGATCTT GGCTCACTGC AACCTCCGCC
1751 TCCTGGGTTA AAGTGATTCT CTCAGCTCAG TCTCCCAGGT AGCTGGGATT
1801 ACAGGCATGT GCCACCACAC CCCGTTAATT TTTGTATTTT TAGTAGAGAC
1851 GGGGTTTCAC CATGTTGGCC AGGCTGGTCT CAAATTCTTG ATCTCAAGTG
1901 ATCTGTCCGC CCCGGCCTCC CAGAGTGCTG GGTTGGGATT ACAGGCGTGA
1951 GCCACCGCGT CCGGCTCAGG ACCCAGTTTT GGCTGCTGGT TCCCAGCAGG
2001 GGACTCGGGG GATATACAGT GGCTGCACCA AATTGGAGGT GTGGGTTCCT
2051 CCAACACAAT TTGCTTCTGC CCGTTGTCTT CCTGCCAGCT GGGTTTGGCC
2101 AGGATTTCTC CGTGTGGGGG CTACATGCGA CCCTCTCCCC TCCTCCCTGA
2151 CTTTAGAGGC TGGTGCTGTG TCGGGAGGAA GGTCAGGGCT CCTGAGCAGC
2201 AATAAAGGAC CAGGAAGAGG CCTGAGGTGG AAAAAAAAAA AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 9 bp to 1346 bp; peptide length: 446
Category: similarity to known protein
Classification: unset
Prosite motifs: WD_REPEATS (323-338)

```
  1 MAARKGRRRT CETGEPMEAE SGDTSSEGPA QVYLPGRGPP LREGEELVMD
 51 EEAYVLYHRA QTGAPCLSFD IVRDHLGDNR TELPLTLYLC AGTQAESAQS
101 NRLMMLRMHN LHGTKPPPSE GSDEEEEEED EEDEEERKPQ LELAMVPHYG
151 GINRVRVSWL GEEPVAGVWS EKGQVEVFAL RRLLQVVEEP QALAAFLRDE
201 QAQMKPIFSF AGHMGEGFAL DWSPRVTGRL LTGDCQKNIH LWTPTDGGSW
251 HVDQRPFVGH TRSVEDLQWS PTENTVFASC SADASIRIWD IRAAPSKACM
301 LTTVTAHDGD VNVISWSRRE PFLLSGGDDG ALKIWDLRQF KSGSPVATFK
351 QHVAPVTSVE WHPQDSGVFA ASGADHQITQ WDLAVERDPE AGDVEADPGL
401 ADLPQQLLFV HQGETELKEL HWHPQCPGLL VSTALSGFTI FRTISV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2c17, frame 3

TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein"; Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence., N = 1, Score = 910, P = 2.7e-91

PIR:S53061 hypothetical protein YMR131c - yeast (Saccharomyces cerevisiae), N = 1, Score = 691, P = 4.3e-68

PIR:I49367 retinoblastoma-binding protein mRbAp46 - mouse, N = 1, Score = 338, P = 1.1e-30

PIR:I39181 retinoblastoma-binding protein RbAp46 - human, N = 1, Score = 338, P = 1.1e-30

>TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein"; Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence.
Length = 469

HSPs:

Score = 910 (136.5 bits), Expect = 2.7e-91, P = 2.7e-91
Identities = 195/442 (44%), Positives = 259/442 (58%)

```
Query:   18 EAESGDTSSEGPAQVYLPGRGPPLREGEELVMDEEAYVLYHRAQTGAPCLSFDIVRDHLG 77
            EA S +  S  P +V+ PG    L +GEEL  D  AY   H    G PCLSFDI+ D LG
Sbjct:   18 EASSSEIPSI-PTRVWQPGVDT-LEDGEELQCDPSAYNSLHGFHVGWPCLSFDILGDKLG 75

Query:   78 DNRTELPLTLYLCAGTQAESAQSNRLMMLRMHNLHGTKP---PPSEGSDEEEEEEDEED- 133
             NRTE P TLY+ AGTQAE A  N + + ++ N+ G +    P + G+ E+E+E+DE+D
Sbjct:   76 LNRTEFPHTLYMVAGTQAEKAAHNSIGLFKITNVSGKRRDVVPKTFGNGEDEDEDDEDDS 135

Query:  134 --------EEERKPQLELAMVPHYGGINRVRVSWLGEEPVAGVWSEKGQVEVFALRRLLQ 185
                    E  + P+++   V H+G +NR+R        +   W++ G V+V+ +   L
Sbjct:  136 DSDDDDGDEASKTPNIQVRRVAHHGCVNRIRAMPQNSH-ICVSWADSGHVQVWDMSSHLN 194

Query:  186 VVEEPQALAAFLRDEQAQMKPIFSFAGHMGEGFALDWSPRVTGRLLTGDCQKNIHLWTPT 245
             + E +              P+ +F+GH  EG+A+DWSP   GRLL+GDC+  IHLW P
Sbjct:  195 ALAESETEGKDGTSPVLNQAPLVNFSGHKDEGYAIDWSPATAGRLLSGDCKSMIHLWEPA 254

Query:  246 DGGSWHVDQRPFVGHTRSVEDLQWSPTENTVFASCSADASIRIWDIRAAPSKACMLTTVT 305
             G SW VD  PF GHT SVEDLQWSP E  VFASCS D S+ +WDIR   S A     +
Sbjct:  255 SG-SWAVDPIPFAGHTASVEDLQWSPAEENVFASCSVDGSVAVWDIRLGKSPAL---SFK 310

Query:  306 AHDGDVNVISWSRREPFLL-SGGDDGALKIWDLRQFKSGSPV-ATFKQHVAPVTSVEWHP 363
            AH+ DVNVISW+R    +L SG DDG   I DLR  K G  V A F+ H  P+TS+EW
Sbjct:  311 AHNADVNVISWNRLASCMLASGSDDGTFSIRDLRLIKGGDAVVAHFEYHKHPITSIEWSA 370
```

```
Query:   364 QDSGVFAASGADHQITQWDLAVERDPE------AGDVEADPGLADLPQQLLFVHQGETEL 417
             ++   A +  D+Q+T WDL++E+D E      A    E      DLP QLLFVHQG+ +L
Sbjct:   371 HEASTLAVTSGDNQLTIWDLSLEKDEEEEAEFNAQTKELVNTPQDLPPQLLFVHQGQKDL 430

Query:   418 KELHWHPQCPGLLVSTALSGFTIFRTISV 446
             KELHWH Q PG+++STA  GF I    ++
Sbjct:   431 KELHWHNQIPGMIISTAGDGFNILMPYNI 459
```

Pedant information for DKFZphfbr2_2c17, frame 3

Report for DKFZphfbr2_2c17.3

[LENGTH] 446
[MW] 49447.38
[pI] 4.82
[HOMOL] TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein"; Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence. 1e-90
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YMR131c] 4e-65
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YEL056w] 4e-15
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YEL056w] 4e-15
[FUNCAT] 06.07 protein modification (glycolsylation, acylation, myristylation, palmitylation, farnesylation and processing) [S. cerevisiae, YEL056w] 4e-15
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YBR195c] 2e-13
[FUNCAT] 10.04.09 regulation of g-protein activity [S. cerevisiae, YBR195c] 2e-13
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YBR195c] 2e-13
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YBR195c] 2e-13
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YBR195c] 2e-13
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YPR178w] 1e-11
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-11
[FUNCAT] 06.13 proteolysis [S. cerevisiae, YGL003c] 4e-09
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL003c] 4e-09
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL145c] 5e-09
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c] 5e-09
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 6e-09
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YMR116c] 5e-08
[FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 5e-08
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YLR429w] 3e-07
[FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR142c] 3e-06
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR142c] 3e-06
[FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YDR142c] 3e-06
[FUNCAT] 03.13 meiosis [S. cerevisiae, YLR129w] 4e-06
[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YER107c] 4e-06
[FUNCAT] 03.01 cell growth [S. cerevisiae, YKL021c] 4e-06
[FUNCAT] 04.07 rna transport [S. cerevisiae, YER107c] 4e-06
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 2e-05
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YCR057c] 2e-05
[FUNCAT] 01.01.04 regulation of amino-acid metabolism [S. cerevisiae, YIL046w] 2e-05
[FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YIL046w] 2e-05
[FUNCAT] 04.01.04 rrna processing [S. cerevisiae, YLL011w] 3e-05
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YOR212w] 5e-05
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YOR212w] 5e-05
[FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YOR212w] 5e-05
[BLOCKS] BL00678
[SCOP] d2trcb_ 2.51.3.1.1 Transducin (heterotrimeric G protein), gamm 5e-29
[PIRKW] plasma 6e-07
[PIRKW] duplication 4e-12
[PIRKW] hormone 6e-07
[PIRKW] transmembrane protein 1e-07
[PIRKW] stomach 6e-07
[PIRKW] actin binding 1e-07
[PIRKW] leucine zipper 1e-07
[PIRKW] signal transduction 2e-06
[PIRKW] heterotrimer 2e-06
[PIRKW] peripheral membrane protein 6e-07
[PIRKW] GTP binding 2e-06
[SUPFAM] WD repeat homology 1e-63
[SUPFAM] yeast coatomer complex alpha chain 1e-07
[SUPFAM] GTP-binding regulatory protein beta chain 4e-07
[SUPFAM] PRL1 protein 8e-09

```
[SUPFAM]     MSI1 protein 4e-12
[SUPFAM]     coatomer complex beta' chain 1e-09
[PROSITE]    WD_REPEATS    1
[PFAM]       WD domain, G-beta repeats
[KW]         All_Beta
[KW]         3D
[KW]         LOW_COMPLEXITY    3.14 %

SEQ    MAARKGRRRTCETGEPMEAESGDTSSEGPAQVYLPGRGPPLREGEELVMDEEAYVLYHRA
SEG    ............................................................
1gotB  ............................................................

SEQ    QTGAPCLSFDIVRDHLGDNRTELPLTLYLCAGTQAESAQSNRLMMLRMHNLHGTKPPPSE
SEG    ............................................................
1gotB  ............................................................

SEQ    GSDEEEEEEDEEDEEERKPQLELAMVPHYGGINRVRVSWLGEEPVAGVWSEKGQVEVFAL
SEG    ..xxxxxxxxxxxxxx............................................
1gotB  ............................................................

SEQ    RRLLQVVEEPQALAAFLRDEQAQMKPIFSFAGHMGEGFALDWSPRVTGRLLTGDCQKNIH
SEG    ............................................................
1gotB  ...........................EEECCCCCEEEEEETTT-TCEEEEEEETTTEEE

SEQ    LWTPTDGGSWHVDQRPFVGHTRSVEDLQWSPTENTVFASCSADASIRIWDIRAAPSKACM
SEG    ............................................................
1gotB  EEETTTT----CEEEEEECCCCCEEEEEEETTCE-EEEEETTTEEEEEETTT--TEEEE

SEQ    LTTVTAHDGDVNVISWSRREPFLLSGGDDGALKIWDLRQFKSGSPVATFKQHVAPVTSVE
SEG    ............................................................
1gotB  EECBTTBTCCEEEEEETTTTTEEEEEETTTEEEEEE........................

SEQ    WHPQDSGVFAASGADHQITQWDLAVERDPEAGDVEADPGLADLPQQLLFVHQGETELKEL
SEG    ............................................................
1gotB  ............................................................

SEQ    HWHPQCPGLLVSTALSGFTIFRTISV
SEG    ..........................
1gotB  ..........................
```

Prosite for DKFZphfbr2_2c17.3

```
PS00678    323->338    WD_REPEATS    PDOC00574
```

Pfam for DKFZphfbr2_2c17.3

```
HMM_NAME    WD domain, G-beta repeats

HMM              *MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*
                  ++GH+  V ++ +SP  + +++S S D ++R+WD
Query       257   FVGHTRSVEDLQWSPTENTVFASCSADASIRIWD    290

24.88   304   336    1    34 dkfzphfbr2_2c17.3 similarity to YMR131c and retinoblastoma-
binding protein RbAp46
  Alignment to HMM consensus:
Query            *MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*
                  + H+++V+ +++S    + ++SG++DG +++WD
  dkfzphfbr2  304  VTAHDGDVNVISWSRREPF-LLSGGDDGALKIWD    336
```

DKFZphfbr2_2c18
----

group: brain associated

DKFZphfbr2_2c18 encodes a novel 302 amino acid protein with weak similarity to cyclin-dependent kinase p130-PITSLRE.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to cyclin-dependent kinase p130-PITSLRE

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2835 bp
Poly A stretch at pos. 2817, polyadenylation signal at pos. 2796

```
   1 TGGGGCGGAC GGCGAGGGAG TCCAGAGCCT TGAGCCCGGT GCTCCTCCCT
  51 CGCGCAGCGG TGGCTCTGCG GCCGCTGGAG TAAACACTGC CTTTGTTCCC
 101 TAGCGCCTCG TCTTTCGTCG CCCCGTGCCC TCACGCCGCC GGGCTCTGGC
 151 CGGCCCGCCC TCGGTCCTTG AACCCCATTT CGGCTCGTGC CGTGCGGATG
 201 CAGCTGCCGG GCCTGGGTTT GGGCATTGAG CGGGAGGAGG AGGAGGAGCG
 251 GCGGCGCCTG GGCGGCATGC GATGGGGAAC TGCTGCTGGA CGCAGTGCTT
 301 CGGACTGCTT CGCAAGGAAG CGGGGCGGCT GCAGCGAGTA GGCGGCGGCG
 351 GAGGATCCAA GTATTTTAGA ACATGCTCAA GAGGTGAGCA CTTGACAATA
 401 GAGTTTGAGA ATCTAGTAGA AAGTGATGAA GGGGAGAGCC CAGGAAGCAG
 451 TCATAGGCCT CTTACTGAGG AAGAAATTGT TGACCTAAGA GAAAGGCATT
 501 ATGATTCCAT TGCCGAAAAA CAAAAAGATC TTGATGAGAA AATTCAAAAA
 551 GAGTTAGCCT TACAAGAAGA GAAGTTAAGA CTAGAAGAAG AAGCTTTATA
 601 CGCTGCACAG CGTGAAGCAG CCAGGGCAGC AAAGCAGCGA AAGCTCTTGG
 651 AGCAAGAAAG GCAGAGAATT GTGCAGCAAT ATCATCCTTC CAACAATGGA
 701 GAATATCAAA GTTCAGGACC AGAAGATGAC TTCGAATCTT GTTTGAGAAA
 751 TATGAAGTCA CAGTATGAAG TTTTTCGAAG TAGTAGACTC TCATCAGATG
 801 CTACAGTTTT GACACCAAAT ACAGAAAGCA GTTGTGATTT AATGACCAAA
 851 ACTAAATCAA CTAGTGGAAA TGACGACAGC ACATCCTTAG ATCTAGAGTG
 901 GGAAGATGAA GAAGGAATGA ATAGAATGCT TCCAATGAGA GAACGTTCCA
 951 AAACAGAGGA AGACATTCTA CGGGCAGCAC TTAAGTATAG CAACAAGAAG
1001 ACTGGAAGTA ATCCTACATC AGCCTCTGAT GATTCCAATG GGCTGGAGTG
1051 GGAAAATGAT TTTGTTAGTG CCGAAATGGA TGATAATGGA AATTCCGAGT
1101 ATTCTGGATT TGTAAATCCT GTATTAGAAC TGTCTGATTC TGGCATAAGG
1151 CATTCTGACA CAGATCAACA GACTCGATAG GGTAAAATTG TGTGACCTTG
1201 TTTATCAGTT ATGACCAAAT GTTAAAAACC AACTAGAATG TATAAGTGAT
1251 TGTGCTTAGC CTTTTTGTAA GGGAGATGTG TAAGAAACCA TGCTGTAAAT
1301 GCTTATTTTA TTACAAAGGA GTAGGGATGA TAGGATCTGA ATTGATACAG
1351 AATTAAGTGC AATTTCATCA TCTGCCTTCT GCTTTTCAAG ACCAATTTAA
1401 TGGTCCTGTC ATGTTACTGA TTAAATTTAC TTTGTCTTGT CTTTATAGCA
1451 TTTCTGTTTA CTATGGTAGA TTTCCACTTT CAATTTTTAA AATTAATTTT
1501 ACTTTGAATG ATTTATGAAG CCTATTTCAT TGTCTAACTA TGAAAATATT
1551 AAGACTTTTT TGTTAATTCT CAGCCGATGT GAAGGAAGCA TGAGGAGGGA
1601 TCGTCAGACT CAGATTTAGA ATAGTGTTCC CGTTTCCAGC ATTATTTATT
1651 TCTATGACTT CTTTGGATTT TATTATCTAA TAGTAAGTAC AGTTGATGTG
1701 GGTAGATGAC TCTAAGAAAT GCTGAAGTAT CGGCATTACA TGTGTTTATT
1751 TACATGTCCT AGTTTGATAA TGTTGATTCA ATCTGAACAA AAGATAATAT
1801 AAAAATAACC CTTCAGAGTT TGGACATTTC AAGTTGGTAA TAATAAAAAA
1851 TAATATTTAA GAAGATATAT ATATATATAT ATTTAGTTTT TTCCACTTCA
1901 TTTTACATGC CACTATATTG ACTTTAATTG ATATACAGTA TTAAGTTTTT
1951 AGGTGCCATT ATTTTTAAAA AATTCTATAT TTCCAATGAA CGATGTTAGA
2001 TTTTACACAG AACATATTCT CTGCATGATT TCAGAAAAGA AAATCTAAAA
2051 AGGTAATACG GGTATTTCAA ATAAAATCCT TTCTGGTATG AAAGGCTCCA
2101 TTGATTTTAT TAAGCCTTCC TTTACCTTGT AGTACAAGGT GCTTTAATGG
2151 GATAGAACTA AGCATATCAA TATCTATAAC TGCATTTTGT GCTAGACAAT
2201 TACTGTTCTT TTCTCTAAAA TGTATATGTC AATTTACAAG GCCAGGGATA
2251 GAAAACACTC CATAATTGCT TTCCTTGATT TTGCTGAGGA TTTGGTATGA
2301 TTTTAGTAAG CAAACTGTTT TTTGGTTTTT CCTTAATGTT TTTAATTTTT
2351 TTTCCTCTTG CAACAATGAC GGTGCATGTT CTTATAAATA TAGGAAGGTC
2401 CAGATATAAA TAGTAACCTA AAGTTCTTGC TGTGCTTAAA AAAAAAAATC
2451 ATGTGGCTCT TTCAATATTT GAACTGCTAA GCAATGACAT CTGTAGTTTT
2501 ATCTCCTTTT TTATGTCATA GAAATTAATA TGATACTTTA AATATGTAAA
2551 TATAATACAT TGGTAATGCT ATTATTTATA TCTGTCTTAA CATAATTTAA
2601 GTTGTAGCTG TGTCTTGGAA ATATTTTTAA GGTAATCTAT ATTCACATTG
2651 CCTGTGTTAA TGCTTTTTAA GGTTTGTATA CATCAGATGT ATATTTTTGG
```

```
2701 TTTGGCATAA GCTACGATTG TAATTTTTCT TGGCTTTTTG TTCATAAAGA
2751 ATTTTTTGAA GGAATGGTAA CAAATGGTAA TTTACAAATG GTTGTGAATA
2801 AACACATTTT TACACTTAAA AAAAAAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 272 bp to 1177 bp; peptide length: 302
Category: similarity to known protein

```
  1 MGNCCWTQCF GLLRKEAGRL QRVGGGGGSK YFRTCSRGEH LTIEFENLVE
 51 SDEGESPGSS HRPLTEEEIV DLRERHYDSI AEKQKDLDEK IQKELALQEE
101 KLRLEEEALY AAQREAARAA KQRKLLEQER QRIVQQYHPS NNGEYQSSGP
151 EDDFESCLRN MKSQYEVFRS SRLSSDATVL TPNTESSCDL MTKTKSTSGN
201 DDSTSLDLEW EDEEGMNRML PMRERSKTEE DILRAALKYS NKKTGSNPTS
251 ASDDSNGLEW ENDFVSAEMD DNGNSEYSGF VNPVLELSDS GIRHSDTDQQ
301 TR
```

BLASTP hits

Entry A55817 from database PIR:
cyclin-dependent kinase p130-PITSLRE - mouse
Length = 783
Score = 123 (43.3 bits), Expect = 0.00013, P = 0.00013
Identities = 53/197 (26%), Positives = 96/197 (48%)

Alert BLASTP hits for DKFZphfbr2_2c18, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2c18, frame 2

Report for DKFZphfbr2_2c18.2

```
[LENGTH]    302
[MW]        34281.39
[pI]        4.73
[PROSITE]   MYRISTYL           5
[PROSITE]   CK2_PHOSPHO_SITE   12
[PROSITE]   TYR_PHOSPHO_SITE   2
[PROSITE]   PKC_PHOSPHO_SITE   3
[KW]        All_Alpha
[KW]        LOW_COMPLEXITY     13.58 %
[KW]        COILED_COIL        13.58 %
```

```
SEQ    MGNCCWTQCFGLLRKEAGRLQRVGGGGGSKYFRTCSRGEHLTIEFENLVESDEGESPGSS
SEG    ......................xxxxx.................................
PRD    cccccccchhhhhhhhhheeeccccccccceeeecccccchhhhhhhhcccccccccc
COILS  ............................................................

SEQ    HRPLTEEEIVDLRERHYDSIAEKQKDLDEKIQKELALQEEKLRLEEEALYAAQREAARAA
SEG    ..............................xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD    ccchhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  ...........................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ    KQRKLLEQERQRIVQQYHPSNNGEYQSSGPEDDFESCLRNMKSQYEVFRSSRLSSDATVL
SEG    xxxxxxx.....................................................
PRD    hhhhhhhhhhhhhhhhcccccccccccccccchhhhhhhhhhheeeeeccccccceeee
COILS  CCCCCCCCC...................................................
```

```
SEQ    TPNTESSCDLMTKTKSTSGNDDSTSLDLEWEDEEGMNRMLPMRERSKTEEDILRAALKYS
SEG    ............................................................
PRD    cccccccccccccccccccccccccchhhhhhhcccccchhhhhhhcchhhhhhhhhhhc
COILS  ............................................................

SEQ    NKKTGSNPTSASDDSNGLEWENDFVSAEMDDNGNSEYSGFVNPVLELSDSGIRHSDTDQQ
SEG    ............................................................
PRD    cccccccccccccccccccccccceeeccccccccccccccceeeeccccccccccccccc
COILS  ............................................................

SEQ    TR
SEG    ..
PRD    cc
COILS  ..
```

Prosite for DKFZphfbr2_2c18.2

PS00005	60->63	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	240->243	PKC_PHOSPHO_SITE	PDOC00005
PS00006	36->40	CK2_PHOSPHO_SITE	PDOC00006
PS00006	65->69	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	163->167	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	198->202	CK2_PHOSPHO_SITE	PDOC00006
PS00006	204->208	CK2_PHOSPHO_SITE	PDOC00006
PS00006	226->230	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	295->299	CK2_PHOSPHO_SITE	PDOC00006
PS00007	103->111	TYR_PHOSPHO_SITE	PDOC00007
PS00007	103->111	TYR_PHOSPHO_SITE	PDOC00007
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	199->205	MYRISTYL	PDOC00008
PS00008	245->251	MYRISTYL	PDOC00008
PS00008	291->297	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_2c18.2)

DKFZphfbr2_2d15
----

group: differentiation/development

DKFZphfbr2_2d15 encodes a novel 438 amino acid protein similarity to Mus musculus testis-specific Y-encoded-like protein (Tspyl1).

The TSPY genes are arranged in clusters on the Y chromosome of many mammalian species. TSPY is believed to function in early spermatogenesis and is a candidate for GBY, the putative gonadoblastoma-inducing gene on the Y. The novel protein is a new member of the TSPY-SET-NAP1L1 family, which represents proteins closely related to TSPY. Therefore, the new protein seems to be involved in early spermatogenesis.

The new protein can find application in modulating early spermatogenesis.

strong similarity to testis-specific Y-encoded-like protein

complete cDNA, complete cds, EST hits
localisation: primer B does not match perfect

Sequenced by Qiagen

Locus: /map="729.2 cR from top of Chr6 linkage group"

Insert length: 3229 bp
Poly A stretch at pos. 3206, polyadenylation signal at pos. 3184

```
   1 GGAGACTGTA GGGTGGGCGG TGCGAGCGGC GGTTAGCTCC CAGTTCGGCC
  51 TCTGAGGAAA ACGGGCGTTC GCCTGCGGTT GGTCCGACTG TTAGCAACAT
 101 GAGCGGCCTG GATGGGGTCA AGAGGACCAC TCCCCTCCAA ACCCACAGCA
 151 TCATTATTTC TGACCAAGTC CCGAGCGACC AGGACGCACA CCAGTACCTG
 201 AGGCTCCGCG ACCAAAGCGA GGCGACACAG GTGATGGCGG AGCCGGGTGA
 251 GGGAGGCTCG GAGACCGTCG CGCTCCCGCC TTCACCGCCT TCAGAGGAGG
 301 GGGGCGTACC CCAGGATCCC GCGGGCCGTG GCGGTACTCC CCAGATCCGA
 351 GTTGTTGGGG GTCGCGGTCA TGTGGCGATC AAAGCCGGGC AGGAAGAGGG
 401 CCAGCCTCCC GCCGAAGGCC TGGCAGCCGC TTCTGTGGTG ATGGCAGCCG
 451 ACCGCAGCCT GAAAAAGGGC GTTCAGGGTG GAGAGAAGGC CCTAGAAATC
 501 TGTGGCGCCC AGAGATCCGC GTCTGAGCTG ACGGCGGGGG CGGAGGCTGA
 551 GGCGGAGGAG GTGAAGACAG GAAAGTGCGC CACCGTCTCA GCAGCCGTGG
 601 CTGAGAGGGA GAGCGCTGAG GTGGTGGTGA AGGAAGGCCT GGCGGAGAAG
 651 GAGGTAATGG AGGAGCAGAT GGAGGTAGAG GAGCAGCCGC CAGAAGGTGA
 701 AGAAATAGAA GTGGCGGAGG AGGATAGATT GGAGGAGGAG GCGAGGGAGG
 751 AAGAAGGGCC CTGGCCTTTG CATGAGGCTC TCCGCATGGA CCCTCTGGAG
 801 GCCATCCAGC TGGAACTGGA CACTGTGAAT GCTCAGGCCG ACAGGGCCTT
 851 CCAACAGCTG GAGCACAAGT TTGGGCGGAT GCGTCGACAC TACCTGGAGC
 901 GGAGGAACTA CATCATTCAG AATATCCCGG GCTTCTGGAT GACTGCTTTT
 951 CGAAACCACC CCCAGTTGTC CGCCATGATT AGGGGCCAAG ATGCAGAGAT
1001 GTTAAGGTAC ATAACCAATT TAGAGGTGAA GGAACTCAGA CACCCTAGAA
1051 CCGGTTGCAA GTTCAAGTTC TTCTTTAGAA GAAACCCCTA CTTCAGAAAC
1101 AAGCTGATTG TCAAGGAATA TGAGGTAAGA TCCTCCGGCC GAGTGGTGTC
1151 TCTTTCTACT CCAATTATAT GGCGCAGGGG GCATGAACCC CAGTCCTTCA
1201 TTCGCAGAAA CCAAGACCTC ATCTGCAGCT TCTTCACTTG GTTTTCAGAC
1251 CACAGCCTTC CAGAGTCCGA CAAAATTGCT GAGATTATTA AAGAGGATCT
1301 GTGGCCAAAT CCACTGCAAT ACTACCTGTT GCGTGAAGGA GTCCGTAGAG
1351 CCCGACGTCG CCCGCTAAGG GAGCCTGTAG AGATCCCCAG GCCCTTTGGG
1401 TTCCAGTCTG GTTAACATTT GCCCTTGGGA ATACTCCTGC ACAAGGTCTC
1451 CTACCACCTT CTGCTGGACC TGTGCTTGGG CATCAGCAAT GAGTATGCCT
1501 TCTATTGTGC TTTGTTTTTG CTGACTTTTC TGCACCCTGT TTCCTTTGGA
1551 TATTCAGTTC TCTCAACCTC AAGATTGAGA CGGTGGTGGG TATGCTTCTC
1601 CACTTCCATA TGACCTTCAT GCTGTTCTGG AATATCACAT GCTACGAGGT
1651 CATCCTTCAC ACTACTTGTA AGCCAAGCAA ATGATACTGT AGATTGTACT
1701 GCCTTTATCT GCACTGCTTG GACCCTGTTT ATTCCCAGGG CCTCTGAACT
1751 GGTTGCTGTC ACTTGGATTT CTAGCTTTGG GAGCCTGTTC CACCTACTCA
1801 GCTCTGCATT GAGCAGTATG GGCACATGCC CTGTGGACAG TTACTGGACG
1851 TTAATGAACT CAGAGGAGAA AAGCAGTGAG CCACTTGTTC TGTGTGATTT
1901 ATGGTACTTC ATTGCTCTTC CTTCACCTCT AGTCACTTTC TATTGCTACC
1951 TGCCCTACAT TGGCTCCTGC CAAGGTCCCT CTCTCTCCCT GTTTTCCTTT
2001 TTTTTTTTTT TTTTTTTTTT TTTTGAGACG GAGGACGGAG TCTTGCTCTG
2051 TCGCCCAGGT TGGAGTGCAG TGGCGCGATC TCGGCTCACT GCAACCTCCA
2101 CCTCCCGGGT TCAAGCGATT CTCCTGCCTC AGCCTCCCGA GTAGCTGGGA
2151 CTACAGGCGC GCGCCGCCAC GCCCGGCTAA TTTTTATATT TTTAGTAGAG
2201 ACGGGGTTTC ACCATGCTGG CCAGGCTGGT CTCGAACCCC GACCTCGTGA
2251 TCCGCCCTCC TTAGCCTCCC AATCCTCTCT TAAAAAAGTG ATAGCTCAGA
2301 AATATTTGTA AAAGCAAGGT TTTTATTTCA TTTTGGCTCT GTCATTTTCA
2351 GAGGCAAAGA AGTTGGCCTG TAAAATAGAG TGCTAGAGCT CTTACGCCCC
2401 TCCCCTTCTT CCCAACTTCC TACTTCCTAG CCCTTTTATC AACTCCTAGA
2451 ATAGTTAAAG AGAGACACAT CTAGATGGGA TGAAAGGTGC CCTAAGCAGG
```

```
2501 AGAAACTGAA CAAAAGGCTA GAGGCATGGG CCAGGTAAAA ATTGGGCCTA
2551 GAGTGAAGAC TGTGCTGCCG TTAAGAGCTT TCGAGGAAGG AGTACTTACT
2601 CCCCAATGAT GATGAATGGA GAAATACTTT TCAGGGAGAA TTGAAGGGGT
2651 TAAAGTGTTA AATATGTTGC CTAGACAAGG GTTCTTTAAA GAAAGACAGC
2701 GCAACTTTGA ATGCTTTCTT ACTTGTTTTG TGACCTAATT TATGTGGAAG
2751 ATTGTTATTT CATTAGGATT TAGTAAAATT TTTTTTTCTG ATTCTAAACT
2801 TATTGTGAAA ATTGAGCTGT ACAGATATTC TTTTGATTTC AATTGGGAAC
2851 ATTTGGAAGA ACAACAGTCT TACTTGCCTG TACAATATAG AGACATATGA
2901 ATAGTCATAA CAGTTTTCAA CTTGTTCTTG TTTCTGTTAA ACTATATTCC
2951 TAGAAACATA GTTTGAACAA CTTGGTCTTT GTTAGGCTTG TCAAATTGCC
3001 TTCATGGAAA AATAATCTAC AAAAGTATGG TTTAATTGAT TGTCTTACAT
3051 GATAATTTTC CCTGGCAACA ACTTAGTAAG TGATATATCT TTTTTCCTAA
3101 ATTGCTTAAA TACTGTGAAA TTGCTCTGAC AAATTGGAAG TGTACCATTG
3151 GCATATTTGT CTTCCTTTTT ATGCATGATG GTAAAATAAA AGCATGTTGT
3201 TCTGCTAAGA AAAAAAAAAA AAAAAAAAA
```

BLAST Results

Entry AF042181 from database EMBLNEW:
Homo sapiens testis-specific Y-encoded-like protein (TSPYL) mRNA, partial cds.
Score = 3411, P = 6.9e-148, identities = 685/687

Entry HS938343 from database EMBL:
human STS WI-11947.
Score = 1195, P = 2.1e-46, identities = 273/299

Medline entries

98399864:
Murine and human TSPYL genes: novel members of the TSPY-SET-NAP1L1 family

Peptide information for frame 3

ORF from 99 bp to 1412 bp; peptide length: 438
Category: strong similarity to known protein
Classification: Differentiation/Development

```
  1 MSGLDGVKRT TPLQTHSIII SDQVPSDQDA HQYLRLRDQS EATQVMAEPG
 51 EGGSETVALP PSPPSEEGGV PQDPAGRGGT PQIRVVGGRG HVAIKAGQEE
101 GQPPAEGLAA ASVVMAADRS LKKGVQGGEK ALEICGAQRS ASELTAGAEA
151 EAEEVKTGKC ATVSAAVAER ESAEVVVKEG LAEKEVMEEQ MEVEEQPPEG
201 EEIEVAEEDR LEEEAREEEG PWPLHEALRM DPLEAIQLEL DTVNAQADRA
251 FQQLEHKFGR MRRHYLERRN YIIQNIPGFW MTAFRNHPQL SAMIRGQDAE
301 MLRYITNLEV KELRHPRTGC KFKFFFRRNP YFRNKLIVKE YEVRSSGRVV
351 SLSTPIIWRR GHEPQSFIRR NQDLICSFFT WFSDHSLPES DKIAEIIKED
401 LWPNPLQYYL LREGVRRARR RPLREPVEIP RPFGFQSG
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2d15, frame 3

TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds., N = 1, Score = 1202, P = 3.1e-122

TREMBL:AB018264_1 gene: "KIAA0721"; product: "KIAA0721 protein"; Homo sapiens mRNA for KIAA0721 protein, partial cds., N = 1, Score = 798, P = 2e-79

TREMBL:AB015345_1 gene: "HRIHFB2216"; Homo sapiens HRIHFB2216 mRNA, partial cds., N = 1, Score = 570, P = 2.9e-55

>TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds.
Length = 379

HSPs:

```
 Score = 1202 (180.3 bits), Expect = 3.1e-122, P = 3.1e-122
 Identities = 258/377 (68%), Positives = 283/377 (75%)

Query:    62 SPPSEEGGVPQDPAGR------GGTPQIRVVGGRGHVAIKAGQEE--GQP-P--AEGLAA 110
             SP  +EG     D  G        GTP  R + G         G+     G P P     EGL
Sbjct:     3 SPERDEGTPVPDSRGHCDADTVSGTPDRRPLLGEEKAVTGEGRAGIVGSPAPRDVEGLVP 62

Query:   111 ASVVMAADRSLKK-GVQGGEKALEICGAQRSASELTAGAEAEAEEVKTGKCATVSAAVAE 169
                 V AA +        V+G    A++ +         ++ T GAE++A +VKT +  TV+AA
Sbjct:    63 QIRVAAARQGESPPSVRGPAAAVFVTPKYVEKAQETRGAESQARDVKT-EPGTVAAAA-- 119

Query:   170 RESAEVVVKEGLAEKEVMEEQMEVEEQPPEGEEIEVAEEDRLEEEAREEEGPWPLHEALR 229
              E +EV                EE MEVE Q P GEE+E+ E         EA EE GPW L    LR
Sbjct:   120 -EKSEVATPGS-------EEVMEVE-QKPAGEEMEMLEASGGVREAPEEAGPWHLGIDLR 170

Query:   230 MDPLEAIQLELDTVNAQADRAFQQLEHKFGRMRRHYLERRNYIIQNIPGFWMTAFRNHPQ 289
              +PLEAIQLELDTVNAQADRAFQ LE KFGRMRRHYLERRNYIIQNIPGFWMTAFRNHPQ
Sbjct:   171 RNPLEAIQLELDTVNAQADRAFQHLEQKFGRMRRHYLERRNYIIQNIPGFWMTAFRNHPQ 230

Query:   290 LSAMIRGQDAEMLRYITNLEVKELRHPRTGCKFKFFFRRNPYFRNKLIVKEYEVRSSGRV 349
             LSAMIRG+DAEMLRY+T+LEVKELRHP+TGCKFKFFFRRNPYFRNKLIVKEYEVRSSGRV
Sbjct:   231 LSAMIRGRDAEMLRYVTSLEVKELRHPKTGCKFKFFFRRNPYFRNKLIVKEYEVRSSGRV 290

Query:   350 VSLSTPIIWRRGHEPQSFIRRNQDLICSFFTWFSDHSLPESDKIAEIIKEDLWPNPLQYY 409
             VSLSTPIIWRRGHEPQSFIRRNQDLICSFFTWFSDHSLPESD+IAEIIKEDLWPNPLQYY
Sbjct:   291 VSLSTPIIWRRGHEPQSFIRRNQDLICSFFTWFSDHSLPESDRIAEIIKEDLWPNPLQYY 350

Query:   410 LLREGVRRARRRPLREPVEIPRPFGFQSG 438
             L REG+RR RRRP+REPVEIPRPFGFQSG
Sbjct:   351 LCREGIRRPRRRPIREPVEIPRPFGFQSG 379
```

Pedant information for DKFZphfbr2_2d15, frame 3

Report for DKFZphfbr2_2d15.3

```
[LENGTH]      438
[MW]          49307.65
[pI]          5.36
[HOMOL]       TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like
protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds. 1e-
107
[FUNCAT]      06.10 assembly of protein complexes  [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]      03.22 cell cycle control and mitosis  [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]      03.04 budding, cell polarity and filament formation  [S. cerevisiae, YKR048c]
1e-07
[FUNCAT]      09.13 biogenesis of chromosome structure    [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]      30.10 nuclear organization    [S. cerevisiae, YKR048c] 1e-07
[BLOCKS]      BL00376F
[PIRKW]       nucleus 6e-39
[PIRKW]       DNA binding 3e-06
[PIRKW]       phosphoprotein 6e-39
[PIRKW]       alternative splicing 6e-39
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY    22.83 %

SEQ    MSGLDGVKRTTPLQTHSIIISDQVPSDQDAHQYLRLRDQSEATQVMAEPGEGGSETVALP
SEG    ...........................................................x
PRD    cccccccccccccceeeeecccccccccchhhhhhhhchhhhhcccccccccceeeecc

SEQ    PSPPSEEGGVPQDPAGRGGTPQIRVVGGRGHVAIKAGQEEGQPPAEGLAAASVVMAADRS
SEG    xxxxxxxxx...................................................
PRD    cccccccccccccccccccceeeeeccccceeeeeccccccccccchhhhhhhhhhhhhcc

SEQ    LKKGVQGGEKALEICGAQRSASELTAGAEAEAEEVKTGKCATVSAAVAERESAEVVVKEG
SEG    ...................xxxxxxxxxxxxx......xxxxxxxxxxxxxxxxxxx.
PRD    cccccccccceeecccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ    LAEKEVMEEQMEVEEQPPEGEEIEVAEEDRLEEEAREEEGPWPLHEALRMDPLEAIQLEL
SEG    .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....................
PRD    hhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhcccccccchhhhhhhhhhhhhh

SEQ    DTVNAQADRAFQQLEHKFGRMRRHYLERRNYIIQNIPGFWMTAFRNHPQLSAMIRGQDAE
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeeccccccccccccccchhh

SEQ    MLRYITNLEVKELRHPRTGCKFKFFFRRNPYFRNKLIVKEYEVRSSGRVVSLSTPIIWRR
```

```
SEG    ............................................................
PRD    hhhhhhhhhhhccccccceeeeeeeecccccchhhhhhcccccccccccccccceeeecc

SEQ    GHEPQSFIRRNQDLICSFFTWFSDHSLPESDKIAEIIKEDLWPNPLQYYLLREGVRRARR
SEG    .................................................xxxxxxxxxxx
PRD    ccccchhhhhcccccceeeeecccccccccchhhhhhhhcccccceeeecccccchhhhh

SEQ    RPLREPVEIPRPFGFQSG
SEG    xxxxxxxx..........
PRD    hccccccccccccccccc
```

(No Prosite data available for DKFZphfbr2_2d15.3)

(No Pfam data available for DKFZphfbr2_2d15.3)

DKFZphfbr2_2d17
----

group: transmembrane proteins

DKFZphfbr2_2d17 encodes a novel 292 amino acid protein with similarity to a C.elegans hypothetical protein.

One transmembrane region is predicted for the protein.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans hypothetical protein

TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 1009 bp
Poly A stretch at pos. 990, polyadenylation signal at pos. 969

```
   1 TGGGCCTGTG GCTGGGGGCA GAGCTCAGAC TGTCTTCTGA AGATTGATGT
  51 CTATTTCCTT GAGCTCTTTA ATTTTGTTGC CAATTTGGAT AAACATGGCA
 101 CAAATCCAGC AGGGAGGTCC AGATGAAAAA GAAAAGACTA CCGCACTGAA
 151 AGATTTATTA TCTAGGATAG ATTTGGATGA ACTAATGAAA AAAGATGAAC
 201 CGCCTCTTGA TTTTCCTGAT ACCCTGGAAG GATTTGAATA TGCTTTTAAT
 251 GAAAAGGGAC AGTTAAGACA CATAAAAACT GGGGAACCAT TTGTTTTTAA
 301 CTACCGGGAA GATTTACACA GATGGAACCA GAAAAGATAC GAGGCTCTAG
 351 GAGAGATCAT CACGAAGTAT GTATATGAGC TCCTGGAAAA GGATTGTAAT
 401 TTGAAAAAAG TATCTATTCC AGTAGATGCC ACTGAGAGTG AACCAAAGAG
 451 TTTTATCTTT ATGAGTGAGG ATGCTTTGAC AAATCCACAG AAACTGATGG
 501 TTTTAATTCA TGGTAGTGGT GTTGTCAGGG CAGGGCAGTG GGCTAGAAGA
 551 CTTATTATAA ATGAAGATCT GGACAGTGGC ACACAGATAC CGTTTATTAA
 601 AAGAGCTGTG GCTGAAGGAT ATGGAGTAAT AGTACTAAAT CCCAATGAAA
 651 ACTATATTGA AGTAGAAAAG CCGAAGATAC ACGTACAGTC ATCATCTGAT
 701 AGTTCAGATG AACCAGCAGA AAAACGGGAA AGAAAAGATA AAGTTTCTAA
 751 AGTAACAAAG AAGCGACGTG ATTTCTATGA GAAGTATCGT AACCCCCAAA
 801 GAGAAAAAGA AATGATGCAA TTGTATATCA GAGTGAGTGA GATCACTACT
 851 TTCCTTTACT ATTTTCTTTA CCTTGTATAT ATTTTATTAT ATGTAGATTG
 901 TTTTGTTTTT CTTCAAGAAT ATTAATTTCT TTATTTGTCA TCATTTATTT
 951 CCCATGGTCG TCTACTTGGA TTAAATGGGT TTTTAAATTC AAAAAAAAAA
1001 AAAAAAAAA
```

BLAST Results

Entry I89937 from database EMBL:
Sequence 11 from patent US 5723315.
Score = 1083, P = 2.2e-42, identities = 223/231

Entry I89938 from database EMBL:
Sequence 12 from patent US 5723315.
Score = 875, P = 7.4e-33, identities = 175/175

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 47 bp to 922 bp; peptide length: 292
Category: similarity to unknown protein
Classification: unset

```
1 MSISLSSLIL LPIWINMAQI QQGGPDEKEK TTALKDLLSR IDLDELMKKD
```

```
 51 EPPLDFPDTL EGFEYAFNEK GQLRHIKTGE PFVFNYREDL HRWNQKRYEA
101 LGEIITKYVY ELLEKDCNLK KVSIPVDATE SEPKSFIFMS EDALTNPQKL
151 MVLIHGSGVV RAGQWARRLI INEDLDSGTQ IPFIKRAVAE GYGVIVLNPN
201 ENYIEVEKPK IHVQSSSDSS DEPAEKRERK DKVSKVTKKR RDFYEKYRNP
251 QREKEMMQLY IRVSEITTFL YYFLYLVYIL LYVDCFVFLQ EY
```

BLASTP hits

Entry S67436 from database PIR:
hypothetical protein - fission yeast (Schizosaccharomyces pombe)
Length = 266
Score = 112 (39.4 bits), Expect = 0.00037, P = 0.00037
Identities = 33/147 (22%), Positives = 69/147 (46%)

Entry CEY75B8A_12 from database TREMBLNEW:
gene: "Y75B8A.31"; Caenorhabditis elegans cosmid Y75B8A
Score = 327, P = 1.5e-29, identities = 72/140, positives = 93/140

Alert BLASTP hits for DKFZphfbr2_2d17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2d17, frame 2

Report for DKFZphfbr2_2d17.2

```
[LENGTH]        292
[MW]            34260.50
[pI]            5.50
[HOMOL]         TREMBLNEW:AF064782_1 product: "unknown"; Mus musculus clone pEN87 unknown mRNA,
partial cds. 1e-119
[KW]            SIGNAL_PEPTIDE 19
[KW]            TRANSMEMBRANE  1
[KW]            LOW_COMPLEXITY    10.96 %

SEQ     MSISLSSLILLPIWINMAQIQQGGPDEKEKTTALKDLLSRIDLDELMKKDEPPLDFPDTL
SEG     .xxxxxxxxxxxxxx.............................................
PRD     ccchhhhhhchhhhhhhcccccccccccchhhhhhhhhhhhcchhhhhhcccccccccc
MEM     ............................................................

SEQ     EGFEYAFNEKGQLRHIKTGEPFVFNYREDLHRWNQKRYEALGEIITKYVYELLEKDCNLK
SEG     ............................................................
PRD     hhhhhhcccccceeeecccccceeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhe
MEM     ............................................................

SEQ     KVSIPVDATESEPKSFIFMSEDALTNPQKLMVLIHGSGVVRAGQWARRLIINEDLDSGTQ
SEG     ............................................................
PRD     eeecccccccccccceeeeeecccccccccceeeeeecccccchhhhhcccccccccccccc
MEM     ............................................................

SEQ     IPFIKRAVAEGYGVIVLNPNENYIEVEKPKIHVQSSSDSSDEPAEKRERKDKVSKVTKKR
SEG     ............................................................
PRD     chhhhhhhhcccceeeeecccccceeeeeccceeeecccccccccchhhhhhhhhhhhhhh
MEM     ............................................................

SEQ     RDFYEKYRNPQREKEMMQLYIRVSEITTFLYYFLYLVYILLYVDCFVFLQEY
SEG     ..........................xxxxxxxxxxxxxxxxxx........
PRD     hhhhhhhccccchhhhhhhhhhhhhheeeeehhhhhhhhhhhhheeeeeeeccc
MEM     ...............................MMMMMMMMMMMMMMMMMMMMMM
```

(No Prosite data available for DKFZphfbr2_2d17.2)

(No Pfam data available for DKFZphfbr2_2d17.2)

DKFZphfbr2_2d20

group: brain derived

DKFZphfbr2_2d20 encodes a novel 197 amino acid protein with similarity to Synechocystis sp. P74594 hypothetical32.8 kD protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to Synechocystis sp. (PCC 6803)

complete cDNA, complete cds, EST hits
potential start at bp 67 matches kozak consensus ANCatgG

Sequenced by Qiagen

Locus: unknown

Insert length: 1787 bp
Poly A stretch at pos. 1768, polyadenylation signal at pos. 1743

```
   1 TGGGGCGGCC GCGGCGGGAA CATGGAGGAG CTGCTGAGGC GCGAGCTGGG
  51 CTGCAGCTCT GTCAGGGCCA CGGGCCACTC GGGGGGCGGG TGCATCAGCC
 101 AGGGCCGGAG CTACGACACG GATCAAGGAC GAGTGTTCGT GAAAGTGAAC
 151 CCCAAGGCGG AGGCCAGAAG AATGTTTGAA GGTGAGATGG CAAGTTTAAC
 201 TGCCATCCTG AAAACAAACA CGGTGAAAGT GCCCAAGCCC ATCAAGGTTC
 251 TGGATGCCCC AGGCGGCGGG AGCGTGCTGG TGATGGAGCA CATGGACATG
 301 AGGCATCTGA GCAGTCATGC TGCAAAGCTT GGAGCCCAGC TGGCCGATTT
 351 ACACCTTGAT AACAAGAAGC TTGGAGAGAT GCGCCTGAAG GAGGCGGGCA
 401 CAGTGTGGAG AGGAGGTGGG CAGGAGGAAC GGCCCTTTGT GGCCCGGTTT
 451 GGATTTGACG TGGTGACGTG CTGTGGATAC CTCCCCCAGG TGAATGACTG
 501 GCAGGAGGAC TGGGTCGTGT TCTATGCCCG GCAGCGCATT CAGCCCCAGA
 551 TGGACATGGT GGAGAAGGAG TCTGGGGACA GGGAGGCCCT CCAGCTTTGG
 601 TCTGCTCTGC AGTAAAAGAT CCCTGACCTG TTCCGTGACC TGGAGATCAT
 651 CCCAGCCTTA CTCCACGGGG ACCTCTGGGG TGGAAACGTA GCAGAGGATT
 701 CCTCTGGGCC GGTGATTTTT GACCCAGCTT CTTTCTACGG CCACTCGGAA
 751 TATGAGCTGG CAATAGCTGG CATGTTTGGG GGCTTTAGCA GCTCCTTTTA
 801 CTCCGCCTAC CACGGCAAAA TCCCCAAGGC CCCAGGATTC GAGAAGCGCC
 851 TTCAGTTGTA TCAGCTCTTT CACTACTTGA ACCACTGGAA TCATTTTGGA
 901 TCGGGGTACA GAGGATCCTC CCTGAACATC ATGAGGAATC TGGTCAAGTG
 951 AGCGGGCCTT ACTCTGGAAG GAGGTCTCAG AGGTTTCTCC ACAGTCCTCT
1001 TCTGGGCAAA TTCTTGTTTC TTCACATGCC GGACTAGCTT AAGACCAATG
1051 CAGTAGCTTA TTTCCAAGCC TTGCAAAGTA TATAATATCT AAGAGGAAAG
1101 GTTTTGTCAT CCCAGCGTTG TCCACTTTGT GGGGCTTTGT AGGTAGACGG
1151 AGCCACACTA CAGGCAGGGT ATGAGCAGAG GGATGTATGG AGTGTGGGCG
1201 ACTCTGAGCC TCACTGCTGC TGCAAGGTGG GGAAACTGTA AGTGAACCCC
1251 TGTGGGTGCG GGGGAGGGTA TCCGGTGCGC AGGGAGGTGG CCAGCGCCCC
1301 CGGGCACTGC TGCTCATAGG TACCTTTCCG CTGCCTCCTC CCTGCTCTCC
1351 TGTGCAGGAA TGTCTCTGAG CTGTTCACGT TGATGCTTCT TGGTTGGCAA
1401 GACTTGGGTG TAGACATGAA ACCACCTTAC TAAAAGCGTC TTAAAATGAC
1451 CAATTCCAGA ATCAAGCGTA TTCCGTTTTC CTCCTGCATG ATCCCTGGGC
1501 CCTCCCGCAG GCTGAGCAAG TCTGTAAACT GATTCTGGGA GAAACCAAGC
1551 TGCTGGCCGT AGGATGTCCT TGGGTACATC CAGGAGTCTT CATTGCTTCT
1601 GTTATTACCC CGTCTCCTCT GCCATTTTCT ACAGCTTGCT GAGTTGTCAT
1651 TCCTTTGCAA CATTAAAATA CATGCTGAAC TCATATTTTT CCTTCCTTCA
1701 CTGTTGTAGT AAAGAGACAT ATTTCATGAA TGGCATTGAT GCTAATAAAC
1751 CCTTTGCCCA AAAATTTGAA AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 22 bp to 612 bp; peptide length: 197
Category: similarity to unknown protein
Prosite motifs: LEUCINE_ZIPPER (117-139)

```
  1 MEELLRRELG CSSVRATGHS GGGCISQGRS YDTDQGRVFV KVNPKAEARR
 51 MFEGEMASLT AILKTNTVKV PKPIKVLDAP GGGSVLVMEH MDMRHLSSHA
101 AKLGAQLADL HLDNKKLGEM RLKEAGTVWR GGGQEERPFV ARFGFDVVTC
151 CGYLPQVNDW QEDWVVFYAR QRIQPQMDMV EKESGDREAL QLWSALQ
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2d20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2d20, frame 1

Report for DKFZphfbr2_2d20.1

```
[LENGTH]    197
[MW]        21963.25
[pI]        6.96
[HOMOL]     PIR:S76790 hypothetical protein - Synechocystis sp. (strain PCC 6803) 9e-12
[SUPFAM]    hypothetical protein b1725 1e-06
[PROSITE]   LEUCINE_ZIPPER 1
[PROSITE]   MYRISTYL 2
[PROSITE]   GLYCOSAMINOGLYCAN 1
[PROSITE]   PKC_PHOSPHO_SITE 2
[KW]        Alpha_Beta
```

```
SEQ  MEELLRRELGCSSVRATGHSGGGCISQGRSYDTDQGRVFVKVNPKAEARRMFEGEMASLT
PRD  ccchhhhhccccceeeecccccccceeeccccccccccceeeeeecccchhhhhhhhhhhhh

SEQ  AILKTNTVKVPKPIKVLDAPGGGSVLVMEHMDMRHLSSHAAKLGAQLADLHLDNKKLGEM
PRD  hhhhhheeeeccceeeecccccceeeeecccccccchhhhhhhhhhhhhhhcccccchhh

SEQ  RLKEAGTVWRGGGQEERPFVARFGFDVVTCCGYLPQVNDWQEDWVVFYARQRIQPQMDMV
PRD  hhhhhcccccccccccceeeccccceeeccccccccccccchhhhhhhhhhhhhhhh

SEQ  EKESGDREALQLWSALQ
PRD  hhhccchhhhhhhhccc
```

Prosite for DKFZphfbr2_2d20.1

PS00002	20->24	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	13->16	PKC_PHOSPHO_SITE	PDOC00005
PS00005	67->70	PKC_PHOSPHO_SITE	PDOC00005
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	104->110	MYRISTYL	PDOC00008
PS00029	96->118	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_2d20.1)

DKFZphfbr2_2g18
----

group: brain derived

DKFZphfbr2_2g18 encodes a novel 229 amino acid protein with partial similarity to the humane dJ30M3.2 gene product.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

J30M3.2 extension of genmodel

complete cDNA, complete cds, EST hits
(mouse ESTs with >90% Identities)

Sequenced by Qiagen

Locus: /map="6p22.1-22"

Insert length: 2444 bp
Poly A stretch at pos. 2425, no polyadenylation signal found

```
   1 TGGTCGAGGG TCGACGGTAT CGATAAGTTT TTTTTTTTTT TTTTTTTTTT
  51 TGGAAAGCAA GGATCACACT TCCCCCTCCC TGTTCCTTAA TCCCTTTTCT
 101 AAAAAGGGGG GAAAATCCGG ATGGATTTTA GGGATTGGTC TGGTGTCAGC
 151 TGTGTCTTAT TGCACACCTA AATCCTGATT ATAGGCTTTT CATTTCTCCG
 201 CAAAGCCTTT ATTTTGGCAG TTAAGCCAAA TGTGTTTTCC AGAAAGTTAG
 251 TTATTTTCTC CTCTTTCTTT CCTTTCTTTC CTCCCTTTTT CCCGTCTGAC
 301 CCCAAACGTT ATTGTCCAAA CATGACTGGA CAGCAGCTTT TGTTTCTTGA
 351 CCCTGTAATA TGACAGTCTG CTAATATTGA CAGAAGGTGC AGTTTTTGGG
 401 TTATAGTCGT GATTTTCGCT AATCAATCAT ATTAGCAGGA AAAAAAATGA
 451 CTTGTTTCTG TTGTACTTGA GTCTTAAGAA AAAGTGCCCA TAGTTTAGTG
 501 ACAATTTCCA AAGGCTTTAG TACCACCTGT ATTTCAAAAT GGGGGACCCA
 551 AACTCCCGGA AGAAACAAGC TCTGAACAGA CTACGTGCTC AGCTTAGAAA
 601 GAAAAAAGAA TCTCTAGCTG ACCAGTTTGA CTTCAAGATG TATATTGCCT
 651 TTGTATTCAA GGAGAAGAAG AAAAAGTCAG CACTTTTTGA AGTGTCTGAG
 701 GTTATACCAG TCATGACAAA TAATTATGAA GAAAATATCC TGAAAGGTGT
 751 GCGAGATTCC AGCTATTCCT TGGAAAGTTC CCTAGAGCTT TTACAGAAGG
 801 ATGTGGTACA GCTCCATGCT CCTCGATATC AGTCTATGAG AAGGGATGTA
 851 ATTGGCTGTA CTCAGGAGAT GGATTTCATT CTTTGGCCTC GGAATGATAT
 901 TGAAAAAATC GTCTGTCTCC TGTTTTCTAG GTGGAAAGAA TCTGATGAGC
 951 CTTTTAGGCC TGTTCAGGCC AAATTTGAGT TTCATCATGG TGACTATGAA
1001 AAACAGTTTC TGCATGTACT GAGCCGCAAG GACAAGACTG GAATCGTTGT
1051 CAACAATCCT AACCAGTCAG TGTTTCTCTT CATTGACAGA CAGCACTTGC
1101 AGACTCCAAA AAACAAAGCT ACAATCTTCA AGTTATGCAG CATCTGCCTC
1151 TACCTGCCAC AGGAACAGCT CACCCACTGG GCAGTTGGCA CCATAGAGGA
1201 TCACCTCCGT CCTTATATGC CAGAGTAGAG TACTGACCAG CAAAATGGAG
1251 AAGATCAGAG AATGCAGCAG CAGTTTTTTT TCTTGTTTTC TTACCACTTT
1301 ATTCTTTCAG AGTTTAAAGA AAATGGACTC ATGCACAGAA CACTATGCAT
1351 TTTGAAACTT GTTCATCCTG GATTTTTTTA AATCATTTTT ATCTCAGAAC
1401 TTAAACAAAA ATTAGATGTC GTGCACGGAC TGTGTGAAAG AAGATGCTTT
1451 GCATATTTGC TGCACTGCAT CAGTATCTTA CTAAAAATGT GAAATGAAAG
1501 GACTATTGTA CACTGAAATG CTTAAATGTA TCTGAAAGCA CAAGGTGATA
1551 CTCATTTTTA TGGTCTTCCC ATTTGTGCTG GTTTTTGCCT CTTTGACATC
1601 TGTCATCAGT ATTTAGAGGG TGAGAAGTGA ATGTAACAGG TATAAATAAC
1651 ATTTTTAAAA ACAATAACTT TGCTATAATC ACAGTTGTTC CAGAGCACTG
1701 TCAGATACAT TCTAATGACC AGAACTGGTT TAAAAAAAGA AAATACAACC
1751 ATGGGAAAGA AATCTTAAAT GAAAAACGCA TCTCATTGTA GGCATTTTTG
1801 CCTCATATTT TACTGGGCCA TGTTTGTTTC CTGGTACTCA TGTATTTTTT
1851 TTTTTTCCAG ATCTCTTTCC CCAAGTTGCT ATTGTAAGAG TATTCTGCTG
1901 CGTGTGGATG CAGTTATACA CATTAAAGCA GATCTGGAGT CTGAAGTAGC
1951 TATAAAGCAG CTATAAAACA GAAATACATG CATAGCTGCA GAAACCATGA
2001 TAGGTAGAGG ACTTTTCTTT TGGTTTTGTT TTGTTTTGTT TTGTTTTGTT
2051 TTTGGTTTTA CAGAGAAGAG ATTTTTATTA CAAAGAAAAA AATTCCAGTG
2101 AATTGTGCAG AAATGCTGGT TTTTACACCA TCCTAAAGAA AAACTTTACA
2151 AGGGTGTTTT GGAGTAGAAA AAAGGTTATA AAGTTGGAAT CTTAAATTGT
2201 AAAATTAACC ATTGAGTGTC AAAGTTCTAA AAGCAGAACT CATTTCGTGC
2251 AATGAACATA AGGAAAGACT ACTGTATAGG TTTTTTTTTT TCTCCTTTTA
2301 AATGAAGAAA AGCTTTGCTT AAGGGTTGCA TACTTTTATT GGAGTAAATC
2351 TGAATGATCC TACTCCTTTG GAGTAAGACT AGTGCTTACC AGTTTCCAAT
2401 TGTATTTAGC TTCTGTTGGA ATTTGAAAAA AAAAAAAAAA AAAA
```

BLAST Results

Entry HS338352 from database EMBL:
human STS EST171398.
Score = 1747, P = 3.0e-74, identities = 359/365

Entry HS447255 from database EMBL:
human STS SHGC-10143.
Score = 1717, P = 6.5e-73, identities = 365/383

Entry HS30M3 from database EMBLNEW:
Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands.
Score = 6646, P = 0.0e+00, identities = 1344/1355

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 539 bp to 1225 bp; peptide length: 229
Category: putative protein

```
  1 MGDPNSRKKQ ALNRLRAQLR KKKESLADQF DFKMYIAFVF KEKKKKSALF
 51 EVSEVIPVMT NNYEENILKG VRDSSYSLES SLELLQKDVV QLHAPRYQSM
101 RRDVIGCTQE MDFILWPRND IEKIVCLLFS RWKESDEPFR PVQAKFEFHH
151 GDYEKQFLHV LSRKDKTGIV VNNPNQSVFL FIDRQHLQTP KNKATIFKLC
201 SICLYLPQEQ LTHWAVGTIE DHLRPYMPE
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2g18, frame 2

TREMBLNEW:HS30M3_2 gene: "dJ30M3.2"; product: "dJ30M3.2 (novel protein)"; Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands., N = 1, Score = 470, P = 1.1e-44

>TREMBLNEW:HS30M3_2 gene: "dJ30M3.2"; product: "dJ30M3.2 (novel protein)";
Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands.
Length = 86

HSPs:

Score = 470 (70.5 bits), Expect = 1.1e-44, P = 1.1e-44
Identities = 86/86 (100%), Positives = 86/86 (100%)

```
Query:   144 AKFEFHHGDYEKQFLHVLSRKDKTGIVVNNPNQSVFLFIDRQHLQTPKNKATIFKLCSIC 203
             AKFEFHHGDYEKQFLHVLSRKDKTGIVVNNPNQSVFLFIDRQHLQTPKNKATIFKLCSIC
Sbjct:     1 AKFEFHHGDYEKQFLHVLSRKDKTGIVVNNPNQSVFLFIDRQHLQTPKNKATIFKLCSIC 60

Query:   204 LYLPQEQLTHWAVGTIEDHLRPYMPE 229
             LYLPQEQLTHWAVGTIEDHLRPYMPE
Sbjct:    61 LYLPQEQLTHWAVGTIEDHLRPYMPE 86
```

Pedant information for DKFZphfbr2_2g18, frame 2

Report for DKFZphfbr2_2g18.2

[LENGTH] 229
[MW] 27083.42
[pI] 9.04
[HOMOL] TREMBL:HS30M3_2 gene: "dJ30M3.2"; product: "dJ30M3.2 (novel protein)"; Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands. 6e-47
[PROSITE] MYRISTYL 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 4
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 4
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 5.24 %

```
SEQ  MGDPNSRKKQALNRLRAQLRKKKESLADQFDFKMYIAFVFKEKKKKSALFEVSEVIPVMT
SEG  ............................................................
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhheeeeec

SEQ  NNYEENILKGVRDSSYSLESSLELLQKDVVQLHAPRYQSMRRDVIGCTQEMDFILWPRND
SEG  .............xxxxxxxxxxxx...................................
PRD  cchhhhhhhccccccccccchhhhhhhhhhhhhhccccccccceeecccccceeeecccch

SEQ  IEKIVCLLFSRWKESDEPFRPVQAKFEFHHGDYEKQFLHVLSRKDKTGIVVNNPNQSVFL
SEG  ............................................................
PRD  hhhhhhhhhhhcccccccccccccccccccchhhhhhhhhhhccccceeeeccccceeee

SEQ  FIDRQHLQTPKNKATIFKLCSICLYLPQEQLTHWAVGTIEDHLRPYMPE
SEG  .................................................
PRD  eeeccccccccccceeeeeeeeeeeeecccccccccceeeecccccccccc
```

Prosite for DKFZphfbr2_2g18.2

PS00001	175->179	ASN_GLYCOSYLATION	PDOC00001
PS00004	22->26	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	44->48	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	99->102	PKC_PHOSPHO_SITE	PDOC00005
PS00005	162->165	PKC_PHOSPHO_SITE	PDOC00005
PS00005	189->192	PKC_PHOSPHO_SITE	PDOC00005
PS00006	25->29	CK2_PHOSPHO_SITE	PDOC00006
PS00006	80->84	CK2_PHOSPHO_SITE	PDOC00006
PS00006	162->166	CK2_PHOSPHO_SITE	PDOC00006
PS00006	218->222	CK2_PHOSPHO_SITE	PDOC00006
PS00007	69->77	TYR_PHOSPHO_SITE	PDOC00007
PS00008	70->76	MYRISTYL	PDOC00008
PS00008	168->174	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_2g18.2)

DKFZphfbr2_2h1

group: brain derived

DKFZphfbr2_2h1 encodes a novel 180 amino acid protein with weak similarity to C.elegans D2007.4 protein

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to C.elegans D2007.4 protein

CpG island in 5' region, complete cDNA

Sequenced by Qiagen

Locus: unknown

Insert length: 957 bp
Poly A stretch at pos. 939, polyadenylation signal at pos. 916

```
  1 GGGGGTCCCT GACTTTATAT GGCTGCTCCT GGCGAGCGAC TGAGTCGTCC
 51 GTGAGGAAAA AGAGGCGAGG CTTTTCCGAG ATCGTCTCAG CGATGGCGCT
101 TCGGTCGCGG TTTTGGGGGT TGTTCTCGGT TTGCAGGAAC CCTGGGTGCA
151 GGTTCGCAGC CCTGTCAACC AGCTCCGAGC CGGCAGCGAA ACCTGAAGTG
201 GACCCTGTGG AAAATGAAGC TGTCGCCCCA GAATTCACCA ACCGGAACCC
251 CCGGAACCTG GAGCTTTTGT CTGTAGCCAG GAAAGAGCGG GGCTGGCGGA
301 CGGTGTTTCC CTCCCGTGAG TTCTGGCACA GGTTGCGAGT TATAAGGACT
351 CAGCATCATG TAGAAGCACT TGTGGAGCAT CAGAATGGCA AGGTTGTGGT
401 TTCGGCCTCC ACTCGTGAGT GGGCTATTAA AAAGCACCTT TATAGTACCA
451 GAAATGTGGT GGCTTGTGAG AGTATAGGAC GAGTGCTGGC ACAGAGATGC
501 TTAGAGGCGG GAATCAACTT CATGGTCTAC CAACCAACCC CGTGGGAGGC
551 AGCCTCAGAC TCGATGAAAC GACTACAAAG TGCCATGACA GAAGGTGGTG
601 TGGTTCTACG GGAACCTCAG AGAATCTATG AATAAATGGA AGCATTAATT
651 GTTTTGAACA TGTAAATATA AATCTGTCAG CCACTACAGC CATCAAAAGA
701 GAGCATCTGG AAGAACAGCC AGCTTGGAAG TTTTACAGCA ATAATGTTGC
751 AGTGGAATAT TATTTGTAGT TAAGGTCATC CTCCTCCCCT TTCTGTTTTT
801 TTAAATCAAG AACTACGTTC TGCCCCTCTC TTGGGCTTCA GAAGCATCTA
851 AGAAAAGCAG TCATCAATTA TAATTAACTT TCAAAGGGCA AGTCAGAAGT
901 TGTTTATAAA TTACAAAATA AAGGCATATT ATGAACTCTA AAAAAAAAAA
951 AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 93 bp to 632 bp; peptide length: 180
Category: similarity to known protein
Classification: unset

```
  1 MALRSRFWGL FSVCRNPGCR FAALSTSSEP AAKPEVDPVE NEAVAPEFTN
 51 RNPRNLELLS VARKERGWRT VFPSREFWHR LRVIRTQHHV EALVEHQNGK
101 VVVSASTREW AIKKHLYSTR NVVACESIGR VLAQRCLEAG INFMVYQPTP
151 WEAASDSMKR LQSAMTEGGV VLREPQRIYE
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2h1, frame 3

PIR:S44789 D2007.4 protein - Caenorhabditis elegans, N = 1, Score = 194, P = 2e-15

PIR:JC5753 ribosomal protein L18 - Vibrio proteolyticus, N = 1, Score = 121, P = 1.1e-07

```
>PIR:S44789 D2007.4 protein - Caenorhabditis elegans
            Length = 170

  HSPs:

 Score = 194 (29.1 bits), Expect = 2.0e-15, P = 2.0e-15
 Identities = 51/134 (38%), Positives = 78/134 (58%)

Query:    48 FTNRNPRNLELLSVARKERGWRTVFP--SREFWHRLRVIRTQHHVEA-LVEHQNGKVVVS 104
             F NRNPRN EL+        G++       +R + +++ ++  + H E  LV +Q+G VV+S
Sbjct:     9 FVNRNPRNNELMGRQAPNTGYQFEKDRAARSYIYKVELVEGKSHREGRLVHYQDG-VVIS 67

Query:   105 ASTREWAIKKHLYSTRNVVACESIGRVLAQRCLEAGINFMVYQPTPWEAASDSMKRLQ-- 162
             AST+E +I    LYS  +  A  +IGRVLA RCL++GI+F +   T  EA   S  +
Sbjct:    68 ASTKEPSIASQLYSKTDTSAALNIGRVLALRCLQSGIHFAMPGATK-EAIEKSQHQTHFF 126

Query:   163 SAMTEGGVVLREPQRI 178
              A+ E G+ L+EP  +
Sbjct:   127 KALEEEGLTLKEPAHV 142
```

Pedant information for DKFZphfbr2_2h1, frame 3

Report for DKFZphfbr2_2h1.3

[LENGTH]	180
[MW]	20576.57
[pI]	9.63
[HOMOL]	PIR:S44789 D2007.4 protein - Caenorhabditis elegans 2e-13
[FUNCAT]	j mrna translation and ribosome biogenesis [H. influenzae, HI0794] 2e-04
[SUPFAM]	Escherichia coli ribosomal protein L18 8e-06
[KW]	Alpha_Beta

```
SEQ   MALRSRFWGLFSVCRNPGCRFAALSTSSEPAAKPEVDPVENEAVAPEFTNRNPRNLELLS
PRD   cccccccceeeeeeeecccccceeeecccccccccccccccccceeeecccccccccchhhhh

SEQ   VARKERGWRTVFPSREFWHRLRVIRTQHHVEALVEHQNGKVVVSASTREWAIKKHLYSTR
PRD   hhhhccccccccchhhhhhhhhhcccccchhhhhhhhhccccceeeeechhhhhhhhhhhcc

SEQ   NVVACESIGRVLAQRCLEAGINFMVYQPTPWEAASDSMKRLQSAMTEGGVVLREPQRIYE
PRD   ccceeehhhhhhhhhhhhcceeeeecccccchhhhhhhhhhhhhhcccceeeccccccccc
```

(No Prosite data available for DKFZphfbr2_2h1.3)

(No Pfam data available for DKFZphfbr2_2h1.3)

DKFZphfbr2_2h10

group: brain derived

DKFZphfbr2_2h10 encodes a novel 220 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2176 bp
Poly A stretch at pos. 2161, polyadenylation signal at pos. 2143

```
   1 TGGGGAGTAT TCTAATTATA TTTTATATTT AATAAATTAT TTTTCTATTT
  51 CTTTGTTATA TTAAGTTGCA CACTTGTTTC TTTTATCCAG AAAGTTTAGT
 101 ATAATAAAAA TAGTTTTAAG ATTAACTGTG AATGTAAAGG AAAAGTATTA
 151 TTAATTATTT CAGGAAATTG CAAGACCTAA CATGGCTGAA AGAGAAACAG
 201 AAACATCAAA TTCTGAAAGT AAACAAGATA AAGCTGCTTC TTCAAAAGAA
 251 AAAAATGGAT GTAATGCAAA TTCATTTGAA GGCTCATCAA CAACAAAAAG
 301 TGAAGAAAGC ATAACAGTTT CAGATAAGGA AAATGAAACC TGTCTTGCAG
 351 ACCAGGAAAC TGGCTCAAAA AACATCGTCA GTTGTGATTC AAATATTGGT
 401 GCAGATAAAG TGGAAAAGAA AAAACAAATA CAACACGTTT GTCAGGAAAT
 451 GGAGTTGAAG ATGTGCCAGA GTTCAGAAAA CATAATCTTA TCTGATCAGA
 501 TTAAAGATCA CAACTCCAGT GAAGCCAGAT TTTCTTCAAA GAATATTAAG
 551 GATTTGCGAT TAGCATCAGA TAATGTAAGC ATTGATCAGT TTTTGAGAAA
 601 AAGACATGAA CCTGAATCTG TTAGTTCTGA TGTTAGCGAG CAAGGCAGTA
 651 TTCATTTGGA ACCTCTGACT CCATCCGAGG TACTTGAGTA TGAAGCCACA
 701 GAGATTCTTC AGAAAGGTAG TGGTGATCCT TCAGCCAAGA CTGATGAAGT
 751 AGTGTCTGAT CAAACAGATG ACATTCCTGG AGGAAATAAC CCTAGCACAA
 801 CAGAGGCAAC AGTAGACCTG GAAGATGAAA AAGAAAGAAG TTGAAATTAG
 851 TCATTTTAAG TTTCAGTGTA CCAACGATAA GGGCATTTGG AACAGTGCTA
 901 TCAGGTGAGC TCAGTGGTGC TGTTGTAGGT TCAGAAATGG AAATATGTAA
 951 GGGAGGTCAC ACATACACTT TACCTGTATG TTCAACCTAT GTTATCAAAC
1001 AAACCAATTC ACCAATAATA GCATGATTAG TAGGGATTCC CAAAAAGTTT
1051 TTAAAAACAC GAACAGGATT TTAATGATAA TTAAATTTGC AGTGGAAAGG
1101 TCTCATTTAA TGGTTTTCAA GGAAATGGGA TTTGGTTGCT GACATGAATT
1151 GATGATATTA GTAATATTTA TAAAGCCTTT CAAACTTCCA TCAATCCTAA
1201 GCTAAAAATC TTTATTACCT GTATATCCTT TTCAGTTAAC TGAGAGGAAG
1251 GGATTTGGAA ACCATGTACT TTTGGGGAGT AATTGATTAA AAACAATGGC
1301 TGATTGGCAT TGTTAATGAA GGCTTTATTT GTGAGGATGA TGCTGGTAAA
1351 TGGAGCATGC TTAGAGTACT AAATTGATCT AATGAGAATT TGGATGAACA
1401 TAAACTTAAT TTTGGATTTA ATATAACATT CCAGTCAGAC GCATGTAAAC
1451 AGAATATTTG AATCTTTGTA CCTCCATACA AGTGTTAGCC TGCCAGGCTG
1501 TAAGCTTACC TTAATTAAAC TTTCAGTGAA AGTGGAATTA TTAAGATATA
1551 AATTTATATT TGTGCTTTTT GTCAGTGTGT AAGCTGTGTA GAAATTCTTT
1601 GATGTATTAG TTGTATTAAT GTAAAGTAGA AACCCATTGT TGAAACTCCT
1651 GTAGCTATTA TGCTTTTAAT ATTGTTTTAA TGTTCTTCCT TAGAAATAGG
1701 CCCATAAAAA TGGTCTGGAA GCCAAACCAA AGTATGGTAT AATGTAGATA
1751 TTGTAAAGCA GTAAACTGAA AACATGTCCT GGCATGTATT CAGCCATGTT
1801 TAAGTGACTT TTCTGTAATT GTAAAATAAA AACTTCAAAT GGGACCTAAA
1851 ACAGTGATGT AAAAGAACTG GTTTTGGAAA TTTAGCCTAA TTTATCTATA
1901 AGATGGCTGC TAAATTGATT TTTCAGTTCT TTTTATCATC TAAAATATAA
1951 TAGATATAGA AATGAATAAT ATGAAGAACA GTAGTTTGCT TTGAAATACT
2001 AATAAACTTT TATTTAAGAT GCTTCATTTT TACTTCTTAA AACGTGCTTT
2051 GGATTCTTAA ATTTTGTTTC ACTGAATGTT CAATGTTTTA AATGGCGATT
2101 AAAATACTCT GCTGTATATA GTAGTTTTTG AGTAAATATT TGCAATAAAA
2151 ATCTGCCCCC GAAAAAAAAA AAAAAA
```

BLAST Results

Entry G35287 from database EMBL:
human STS SHGC-37375.
Score = 2163, P = 2.8e-91, identities = 437/441

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 182 bp to 841 bp; peptide length: 220
Category: putative protein

```
  1 MAERETETSN SESKQDKAAS SKEKNGCNAN SFEGSSTTKS EESITVSDKE
 51 NETCLADQET GSKNIVSCDS NIGADKVEKK KQIQHVCQEM ELKMCQSSEN
101 IILSDQIKDH NSSEARFSSK NIKDLRLASD NVSIDQFLRK RHEPESVSSD
151 VSEQGSIHLE PLTPSEVLEY EATEILQKGS GDPSAKTDEV VSDQTDDIPG
201 GNNPSTTEAT VDLEDEKERS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2h10, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2h10, frame 2

Report for DKFZphfbr2_2h10.2

```
[LENGTH]     220
[MW]         24109.02
[pI]         4.51
[FUNCAT]     04.99 other transcription activities [S. cerevisiae, YKR092c] 4e-05
[FUNCAT]     30.10 nuclear organization    [S. cerevisiae, YKR092c] 4e-05
[PROSITE]    MYRISTYL          3
[PROSITE]    CK2_PHOSPHO_SITE     8
[PROSITE]    PKC_PHOSPHO_SITE     5
[PROSITE]    ASN_GLYCOSYLATION    3
[PFAM]       TNFR/NGFR cysteine-rich region
[KW]         Alpha_Beta
```

```
SEQ  MAERETETSNSESKQDKAASSKEKNGCNANSFEGSSTTKSEESITVSDKENETCLADQET
PRD  cccccccccccccchhhhhhhcccccccccccccccccceeeeeeeecccccccccccccc

SEQ  GSKNIVSCDSNIGADKVEKKKQIQHVCQEMELKMCQSSENIILSDQIKDHNSSEARFSSK
PRD  cccceeeeccccccchhhhhhhhhhhhhhhhhhhhhhcccceeeecccccccccccccccc

SEQ  NIKDLRLASDNVSIDQFLRKRHEPESVSSDVSEQGSIHLEPLTPSEVLEYEATEILQKGS
PRD  cchhhhhhccccchhhhhhhhcccccccccccccccceeecccccccchhhhhhhcccccc

SEQ  GDPSAKTDEVVSDQTDDIPGGNNPSTTEATVDLEDEKERS
PRD  ccccccccccccccccccccccccccceeeehhhhhhccc
```

Prosite for DKFZphfbr2_2h10.2

PS00001	51->55	ASN_GLYCOSYLATION	PDOC00001
PS00001	111->115	ASN_GLYCOSYLATION	PDOC00001
PS00001	131->135	ASN_GLYCOSYLATION	PDOC00001
PS00005	20->23	PKC_PHOSPHO_SITE	PDOC00005
PS00005	37->40	PKC_PHOSPHO_SITE	PDOC00005
PS00005	47->50	PKC_PHOSPHO_SITE	PDOC00005
PS00005	118->121	PKC_PHOSPHO_SITE	PDOC00005
PS00005	184->187	PKC_PHOSPHO_SITE	PDOC00005
PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	13->17	CK2_PHOSPHO_SITE	PDOC00006
PS00006	20->24	CK2_PHOSPHO_SITE	PDOC00006
PS00006	38->42	CK2_PHOSPHO_SITE	PDOC00006
PS00006	45->49	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	163->167	CK2_PHOSPHO_SITE	PDOC00006
PS00006	205->209	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008

PS00008	34->40	MYRISTYL	PDOC00008
PS00008	201->207	MYRISTYL	PDOC00008

Pfam for DKFZphfbr2_2h10.2

```
HMM_NAME     TNFR/NGFR cysteine-rich region

HMM             *CpeG.tYtD.WNHvpqClpCtrCePEMGQYMvqPCTwTQNTVC*
                  +E+ T +D +N  ++C       E   G+  + +C+++  +
Query         40  SEESITVSDKEN--ETC--LADQET--GSKNIVSCDSNIGADK      76
```

DKFZphfbr2_2i17
----

group: intracellular transport and trafficking

DKFZphfbr2_2i17.3 encodes a novel 201 amino acid putative GTP-binding protein related to Rab1B.

Rab proteins are members of the Ras superfamily of GTPases. Rab proteins are localised to the cytoplasmic side of organelles and vesicles involved in the secretory(biosynthetic) and endocytotic pathways in eukaryotic cells. Rab proteins direct the targeting and fusion of transport vesicles to their acceptor membranes. Rab1B is essential for the intracellular transport of nascent low density lipoprotein (LDL) receptor. It is discussed as a universal mediator of endoplasmatic reticulum to Golgi transport of membrane glycoproteins in mammalian cells.

The new protein can find clinical application in modulating the transport of glycoproteins inside cells, especially of the LDL receptor.

Medline
96245776: Intracellular transport and maturation of nascent low density lipoprotein receptor is blocked by mutation in the Ras-related GTP-binding protein, RAB1B

strong similarity to rab1

complete cDNA, complete cds, start at 47, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1985 bp
Poly A stretch at pos. 1901, polyadenylation signal at pos. 1859

```
   1 GGGAGCAGAG TCGACTGGGA GCGACCGAGC GGGCCGCCGC CGCCGCCATG
  51 AACCCCGAAT ATGACTACCT GTTTAAGCTG CTTTTGATTG GCGACTCAGG
 101 CGTGGGCAAG TCATGCCTGC TCCTGCGGTT TGCTGATGAC ACGTACACAG
 151 AGAGCTACAT CAGCACCATC GGGGTGGACT TCAAGATCCG AACCATCGAG
 201 CTGGATGGCA AAACTATCAA ACTTCAGATC TGGGACACAG CGGGCCAGGA
 251 ACGGTTCCGG ACCATCACTT CCAGCTACTA CCGGGGGGCT CATGGCATCA
 301 TCGTGGTGTA TGACGTCACT GACCAGGAAT CCTACGCCAA CGTGAAGCAG
 351 TGGCTGCAGG AGATTGACCG CTATGCCAGC GAGAACGTCA ATAAGCTCCT
 401 GGTGGGCAAC AAGAGCGACC TCACCACCAA GAAGGTGGTG GACAACACCA
 451 CAGCCAAGGA GTTTGCAGAC TCTCTGGGCA TCCCCTTCTT GGAGACGAGC
 501 GCCAAGAATG CCACCAATGT CGAGCAGGCG TTCATGACCA TGGCTGCTGA
 551 AATCAAAAAG CGGATGGGGC CTGGAGCAGC CTCTGGGGGC GAGCGGCCCA
 601 ATCTCAAGAT CGACAGCACC CCTGTAAAGC CGGCTGGCGG TGGCTGTTGC
 651 TAGGAGGGGC ACATGGAGTG GGACAGGAGG GGGCACCTTC TCCAGATGAT
 701 GTCCCTGGAG GGGGGAGGAG GTACCTCCCT CTCCCTCTCC TGGGGCATTT
 751 GAGTCTGTGG CTTTGGGGTG TCCTGGGCTC CCCATCTCCT TCTGGCCCAT
 801 CTGCCTGCTG CCCTGAGCCC CGGTTCTGTC AGGGTCCCTA AGGGAGGACA
 851 CTCAGGGCCT GTGGCCAGGC AGGGCGGAGG CCTGCTGTGC AGTTGCCTCT
 901 AGGTGACTTT CCAAGATGCC CCCCTACACA CCTTTCTTTG GAACGAGGGC
 951 TCTTCTGTCG GTGTCCCTCC CACCCCCATG TATGCTGCAC TGGGTTCTCT
1001 CCTTCTTCTT CCTGCTGTCC TGCCCAAGAA CTGAGGGTCT CCCCGGCCTC
1051 TACTGCCCTG GCTGCAGTCA GTGCCCAGGG CGAGGAATGT GGCCAGGGGA
1101 TCCAGGACCT GGGATCCAGG GCCCTGGGCT GGACCTCAGG ACAGGCATGG
1151 AGGCCACAGG GGCCCAGCAG CCCACCCTTT CCTCTCCCCA CTGCCTCCTC
1201 TCCCTTCCTA CACTCCCAGC TCGAGCCGTC CAGCTGCGGT GGGATCTGAG
1251 TATATCTAGG GCGGGTGGGC GGGTAGCAGT GCTGGGCCTG TGTCTTGAGC
1301 CTGGAGGGAG ACTGCTCCTG CCGCCCTCTG CCCTGCCGGA GACAGACCCA
1351 TGCGCTGCCT GCCCACCGTG CCCCTTTGTC CCCATGTCAG GCGGAGGCGG
1401 AAGGCCCACC GTGCCAGAGG CTGGGCACCA GCCTTAACCC TCACTCTGCT
1451 AGCACCTCCT CCCTTTCCCC AAGGTAGCAC ATCTGGCTCA CTCCCCACTC
1501 CGTCTCTGGA GCCCACCAGG GAAGGCCCTC ATCCCCTGCC GCTACTTCTC
1551 TGGGGAATGT GGGTTCCATC CAGGATTGGG GGCCTCTCTG CTCACCCACT
1601 CTGCACCCAG GATCCTAGTC CCCTGCCCTC TGGCACAGCT GCTTCCTGCA
1651 AGAAAGCAAG TCTTTGGTCT CCCTGAGAAG CCATGTCCCT CGTGCTGTCT
1701 CTTGCCTGTC CCACCTGTGC CCTGCCCTCC AGCTTGTATT TAAGTCCCTG
1751 GGCTGCCCCC TTGGGGTGCC CCCCGCTCCC AGGTTCCCCT CTGGTGTCAT
1801 GTCAGGCATT TTGCAAGGAA AAGCCACTTG GGGAAAGATG GAAAAGGACA
1851 AAAAAAATTA ATAAATTTCC ATTGGCCCTC GGGTGAGCTG AGGGTTTTTG
1901 CAAGGAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1951 AAAAAAAAAA AAAAGAAAAA AAAAAAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

91115900:
A family of ras-like GTP-binding proteins expressed in electromotor neurons.

Peptide information for frame 3

ORF from 48 bp to 650 bp; peptide length: 201
Category: strong similarity to known protein

```
  1 MNPEYDYLFK LLLIGDSGVG KSCLLLRFAD DTYTESYIST IGVDFKIRTI
 51 ELDGKTIKLQ IWDTAGQERF RTITSSYYRG AHGIIVVYDV TDQESYANVK
101 QWLQEIDRYA SENVNKLLVG NKSDLTTKKV VDNTTAKEFA DSLGIPFLET
151 SAKNATNVEQ AFMTMAAEIK KRMGPGAASG GERPNLKIDS TPVKPAGGGC
201 C
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2i17, frame 3

SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B., N = 1, Score = 1023, P = 2.7e-103

PIR:S06147 GTP-binding protein rab1B - rat, N = 1, Score = 1013, P = 3.2e-102

SWISSPROT:RAB1_DISOM RAS-RELATED PROTEIN ORAB-1., N = 1, Score = 967, P = 2.4e-97

PIR:TVHUYP GTP-binding protein Rab1 - human, N = 1, Score = 966, P = 3e-97

>SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B.
Length = 201

HSPs:

Score = 1023 (153.5 bits), Expect = 2.7e-103, P = 2.7e-103
Identities = 197/201 (98%), Positives = 199/201 (99%)

```
Query:   1 MNPEYDYLFKLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIELDGKTIKLQ 60
           MNPEYDYLFKLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIELDGKTIKLQ
Sbjct:   1 MNPEYDYLFKLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIELDGKTIKLQ 60

Query:  61 IWDTAGQERFRTITSSYYRGAHGIIVVYDVTDQESYANVKQWLQEIDRYASENVNKLLVG 120
           IWDTAGQERFRT+TSSYYRGAHGIIVVYDVTDQESYANVKQWLQEIDRYASENVNKLLVG
Sbjct:  61 IWDTAGQERFRTVTSSYYRGAHGIIVVYDVTDQESYANVKQWLQEIDRYASENVNKLLVG 120

Query: 121 NKSDLTTKKVVDNTTAKEFADSLGIPFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG 180
           NKSDLTTKKVVDNTTAKEFADSLG+PFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG
Sbjct: 121 NKSDLTTKKVVDNTTAKEFADSLGVPFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG 180

Query: 181 GERPNLKIDSTPVKPAGGGCC 201
           GERPNLKIDSTPVK A GGCC
Sbjct: 181 GERPNLKIDSTPVKSASGGCC 201
```

Pedant information for DKFZphfbr2_2i17, frame 3

Report for DKFZphfbr2_2i17.3

[LENGTH] 201

[MW] 22171.25
[pI] 5.56
[HOMOL] SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B. 1e-112
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YFL038c] 2e-77
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YFL038c] 2e-77
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YFL005w] 4e-57
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 4e-57
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w] 4e-57
[FUNCAT] 08.19 cellular import [S. cerevisiae, YER031c] 8e-46
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YER031c] 8e-46
[FUNCAT] 09.09 biogenesis of intracellular transport vesicles [S. cerevisiae, YGL210w] 1e-44
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YOR089c] 1e-30
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YNL098c] 3e-25
[FUNCAT] 11.01 stress response [S. cerevisiae, YNL098c] 3e-25
[FUNCAT] 03.99 other cell growth, cell division and dna synthesis activities [S. cerevisiae, YNL098c] 3e-25
[FUNCAT] 01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YNL098c] 3e-25
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YNL098c] 3e-25
[FUNCAT] 10.04.07 g-proteins [S. cerevisiae, YNL098c] 3e-25
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 3e-25
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 9e-24
[FUNCAT] 11.10 cell death [S. cerevisiae, YOR101w] 9e-24
[FUNCAT] 04.07 rna transport [S. cerevisiae, YOR185c] 4e-23
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YOR185c] 4e-23
[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YOR185c] 4e-23
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 7e-17
[FUNCAT] 10.02.07 g-proteins [S. cerevisiae, YPR165w] 7e-17
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 1e-16
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YLR229c] 1e-11
[FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YLR229c] 1e-11
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDL192w] 4e-10
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL180c] 9e-09
[FUNCAT] 06.07 protein modification (glycolsylation, acylation, myristylation, palmitylation, farnesylation and processing) [S. cerevisiae, YPL051w] 3e-08
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YAL048c] 5e-05
[BLOCKS] BL01019A ADP-ribosylation factors family proteins
[BLOCKS] BL01115A GTP-binding nuclear protein ran proteins
[SCOP] d1plk__ 3.25.1.3.1 cH-p21 Ras protein [human (Homo sapiens) 2e-41
[SCOP] d1guaa_ 3.25.1.3.10 Rap1A (Human (Homo sapiens) 5e-60
[SCOP] d1rrga_ 3.25.1.3.5 ADP-ribosylation factor 1 (ARF1) [rat (Rattu 2e-30
[SCOP] d1hura_ 3.25.1.3.4 ADP-ribosylation factor 1 (ARF1) [human (Hom 2e-33
[PIRKW] nucleus 1e-21
[PIRKW] membrane trafficking 1e-110
[PIRKW] oncogene 1e-25
[PIRKW] endoplasmic reticulum 1e-105
[PIRKW] phosphoprotein 1e-105
[PIRKW] glycoprotein 3e-25
[PIRKW] prenylated cysteine 1e-110
[PIRKW] signal transduction 4e-23
[PIRKW] transforming protein 1e-105
[PIRKW] purine nucleotide binding 2e-24
[PIRKW] alternative splicing 5e-26
[PIRKW] P-loop 1e-110
[PIRKW] lipoprotein 1e-110
[PIRKW] proto-oncogene 3e-27
[PIRKW] methylated carboxyl end 3e-27
[PIRKW] hydrolase 7e-25
[PIRKW] membrane protein 1e-105
[PIRKW] GTP binding 1e-110
[PIRKW] thiolester bond 5e-76
[PIRKW] Golgi apparatus 1e-105
[SUPFAM] ras transforming protein 1e-110
[PROSITE] ATP_GTP_A 1
[PROSITE] MYRISTYL 2
[PROSITE] CK2_PHOSPHO_SITE 5
[PROSITE] SIGMA54_INTERACT_1 1
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] GLYCOSAMINOGLYCAN 1
[PROSITE] PKC_PHOSPHO_SITE 4
[PROSITE] ASN_GLYCOSYLATION 3
[PFAM] Ras family (contains ATP/GTP binding P-loop)
[KW] Alpha_Beta
[KW] 3D

```
SEQ     MNPEYDYLFKLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIELDGKTIKLQ
221p-   ........EEEEEEETTTTCHHHHHHHHHHCCCCCCCCCTTTEEEE-EEEEETTEEEEEE

SEQ     IWDTAGQERFRTITSSYYRGAHGIIVVYDVTDQESYANVKQWLQEIDRYASENVNKLLVG
221p-   EEECTTTTTTCGGGHHHHHHCCEEEEEEEETTBHHHHHHHHHHHHHHHHHHTTTTCEEEEE

SEQ     NKSDLTTKKVVDNTTAKEFADSLGIPFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG
221p-   ETTTTCCC-CCCHHHHHHHHHHCCCCEEEETTTTTTTHHHHHHHHHHHHHH.........

SEQ     GERPNLKIDSTPVKPAGGGCC
221p-   .....................
```

Prosite for DKFZphfbr2_2i17.3

PS00001	121->125	ASN_GLYCOSYLATION	PDOC00001
PS00001	133->137	ASN_GLYCOSYLATION	PDOC00001
PS00001	154->158	ASN_GLYCOSYLATION	PDOC00001
PS00002	17->21	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	56->59	PKC_PHOSPHO_SITE	PDOC00005
PS00005	126->129	PKC_PHOSPHO_SITE	PDOC00005
PS00005	135->138	PKC_PHOSPHO_SITE	PDOC00005
PS00005	151->154	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	91->95	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	179->183	CK2_PHOSPHO_SITE	PDOC00006
PS00007	27->34	TYR_PHOSPHO_SITE	PDOC00007
PS00008	18->24	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00017	15->23	ATP_GTP_A	PDOC00017
PS00675	11->25	SIGMA54_INTERACT_1	PDOC00579

Pfam for DKFZphfbr2_2i17.3

```
HMM_NAME     Ras family (contains ATP/GTP binding P-loop)

HMM              *KLVLIGDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGKtIK
                  KL+LIGDSGVGKSCLL+RF +++++E+YI+TIGVDF+++TIE+DGKTIK
Query         10  KLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIELDGKTIK    58

HMM              LQIWDTAGQERYRsMRPMYYRGAMGFMLVYDITNRqSFENIrNWweEIrR
                 LQIWDTAGQER+R+++++YYRGA+G+++VYD+T+++S+ N+++W++EI+R
Query         59 LQIWDTAGQERFRTITSSYYRGAHGIIVVYDVTDQESYANVKQWLQEIDR   108

HMM              HCDrDENVPIMLVGNKCDLEDQRQVStEEGQeFAREWGAIPFMETSAKTN
                 +++  ENV ++LVGNK+DL  +++V+  +++EFA+++G IPF+ETSAK++
Query        109 YAS--ENVNKLLVGNKSDLTTKKVVDNTTAKEFADSLG-IPFLETSAKNA   155

HMM              iNVEEAFMEIvReIlqrMqe.q.NqteNinidQpsrnrk...rCCCIM*
                 +NVE+AFM+++ EI++RM+    +++E +N++ +S++ K     +CC
Query        156 TNVEQAFMTMAAEIKKRMGPGAASGGERPNLKIDSTPVKPAGGGCC--    201
```

DKFZphfbr2_2k19
----

group: brain derived

DKFZphfbr2_2k19 encodes a novel 303 amino acid protein with similarity to human KIAA0378 product.

The protein contains a leucine zipper, which can mediate protein-protein-interaction.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to KIAA0378

encoded by the genomic clones HS147M19/HS608E8

Sequenced by Qiagen

Locus: unknown

Insert length: 1931 bp
Poly A stretch at pos. 1866, no polyadenylation signal found

```
   1 GGGGGGGGCG CGCGGTGACA GCGCGGGGTT GGCGGCGTGG GACCCAGGGG
  51 GCGACAGAGG CAGCAGCAGC CCGAGGCCTG AGGAGAGGAG ACCGGCGGCG
 101 GCGGCAATGC TGGAGACCCT TCGCGAGCGG CTGCTGAGCG TGCAGCAGGA
 151 TTTCACCTCC GGGCTGAAGA CTTTAAGTGA CAAGTCAAGA GAAGCAAAAG
 201 TGAAAAGCAA ACCCAGGACT GTTCCATTTT TGCCAAAGTA CTCTGCTGGA
 251 TTAGAATTAC TTAGCAGGTA TGAGGATACA TGGGCTGCAC TTCACAGAAG
 301 AGCCAAAGAC TGTGCAAGTG CTGGAGAGCT GGTGGATAGC GAGGTGGTCA
 351 TGCTTTCTGC GCACTGGGAG AAGAAAAAGA CAAGCCTCGT GGAGCTGCAA
 401 GAGCAGCTCC AGCAGCTCCC AGCTTTAATC GCAGACTTAG AATCCATGAC
 451 AGCAAATCTG ACTCATTTAG AGGCGAGTTT TGAGGAGGTA GAGAACAACC
 501 TGCTGCATCT GGAAGACTTA TGTGGGCAGT GTGAATTAGA AAGATGCAAA
 551 CATATGCAGT CCCAGCAACT GGAGAATTAC AAGAAAAATA AGAGGAAGGA
 601 ACTTGAAACC TTCAAAGCTG AACTAGATGC AGAGCACGCC CAGAAGGTCC
 651 TGGAAATGGA GCACACCCAG CAAATGAAGC TGAAGGAGCG GCAGAAGTTT
 701 TTTGAGGAAG CCTTCCAGCA GGACATGGAG CAGTACCTGT CCACTGGCTA
 751 CCTGCAGATT GCAGAGCGGC GAGAGCCCAT AGGCAGCATG TCATCCATGG
 801 AAGTGAACGT GGACATGCTG GAGCAGATGG TCCTGATGGA CATATCGGAC
 851 CAGGAGGCCC TGGACGTCTT CCTGAACTCT GGAGGAGAAG AGAACACTGT
 901 GCTGTCCCCC GCCTTAGGTA GGGTTGACAA ACTTGCATTA GCTGAACCAG
 951 GGCAGTATCG ATGCCACTCC CCTCCAAAGG TGAGACGTGA GAACCATCTG
1001 CCAGTCACTT ACGCATAAAC CCCCAAGCTC ACAGCCAGCT CCTGGCTCCC
1051 TAACCCCACG GTTCCACACG GCTGTGTGGC AGCTGCAACA GTGGTGTGGT
1101 TCCGTCATGA ATTCTTCTCA AAGATTTGAC ATGCTCCACT CCGGTAACTT
1151 TGGTGAGTTG AGAGCTTTCT TGTTTGTTTT CCCTCCTTTA CCATCCAGAA
1201 ATCCATTTGA GTCTGCTCCT TGTGGTTAAG GACTGGCGTT TGCAGGGAGG
1251 TGCGGACTCT CCTGCGGGGC TCACGGGAAA CTCTTCCCTC TTCGTGCGAC
1301 AGGCATTTAG GGGCGTGCCT GCCATGGGCA AAGCCATGGT GTGTGTTCAG
1351 CTCTTGGCCT GTGTTGTAAA CTTAGTTGCA CTTCAGTTCC TTTCATCCCT
1401 TCACAAAATT TTGTTTCACA TTCATGCAGC AAATATGGGC TGAGGTGCCA
1451 GACCTGTACC TGGGCTTGGT GCGTTTCAAA TTTCAGACCA GTTCTTTGGG
1501 CTGGGTCAAG GCAAAGCTCA GTCGTCCCAG CAGCACCTCA GCCATCTGTA
1551 GAAGGTTCTA CCATTACCAC GGTTTCAGCT TCCTCTAAAC TTCTCACCCG
1601 CTTCTCCTGG CAATCTGTCA GAACGGTGTC ATCCTGGGGA AGAGAAGGAG
1651 CTTGGGTGCA TTTGCCCTCA TCCTGAGAAG GCCAGAATAC TGGAGACCAG
1701 CGTGAACCCT CACCCAGAGT CAGGGGAAGA TTTAGAAACA GTGACACCTG
1751 CATATAGAAT TTTGATTCCT TGAAGAGCCT ATTTAGTTCC ATAAAATTGG
1801 AGAACTGCTG AAGGTCAGTA ATTCCGACTT TCTCAGCAGT GGTGTCTCTG
1851 AATTACTGCA AAGGGTAAAA AAAAAAAAAA AAAAAACTTA TCGATACCGT
1901 CGACCTCGAT GATGATGATG ATGATGTCGA C
```

BLAST Results

Entry HS147M19 from database EMBL:
Homo sapiens DNA sequence from PAC 147M19 on chromosome 6p22.1-22.3.
Contains an unknown gene, ESTs and GSSs.
Score = 5540, P = 4.1e-275, identities = 1114/1120
3 exons 592-1884

Entry HS608E8 from database EMBL:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 608E8
Score = 797, P = 1.2e-78, identities = 161/163

6 exons 1-592

Medline entries

90294724:
The involucrin gene of the gibbon: The middle region shared by the hominoids

Peptide information for frame 2

ORF from 107 bp to 1015 bp; peptide length: 303
Category: similarity to known protein
Classification: unset
Prosite motifs: LEUCINE_ZIPPER (97-119)

```
  1 MLETLRERLL SVQQDFTSGL KTLSDKSREA KVKSKPRTVP FLPKYSAGLE
 51 LLSRYEDTWA ALHRRAKDCA SAGELVDSEV VMLSAHWEKK KTSLVELQEQ
101 LQQLPALIAD LESMTANLTH LEASFEEVEN NLLHLEDLCG QCELERCKHM
151 QSQQLENYKK NKRKELETFK AELDAEHAQK VLEMEHTQQM KLKERQKFFE
201 EAFQQDMEQY LSTGYLQIAE RREPIGSMSS MEVNVDMLEQ MVLMDISDQE
251 ALDVFLNSGG EENTVLSPAL GRVDKLALAE PGQYRCHSPP KVRRENHLPV
301 TYA
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2k19, frame 2

TREMBL:HSAB2376_1 gene: "KIAA0378"; Human mRNA for KIAA0378 gene, partial cds., N = 1, Score = 137, P = 4.8e-06

PIR:I37037 involucrin - common gibbon, N = 1, Score = 124, P = 7.4e-05

PIR:A57013 early endosome antigen 1 - human, N = 1, Score = 128, P = 9.5e-05

>TREMBL:HSAB2376_1 gene: "KIAA0378"; Human mRNA for KIAA0378 gene, partial cds.
Length = 808

HSPs:

Score = 137 (20.6 bits), Expect = 4.8e-06, P = 4.8e-06
Identities = 59/222 (26%), Positives = 103/222 (46%)

```
Query:     2 LETLRERLLSVQQDFTSGLKTL---SDKSREAKVKS-KPRTVPFLPKYSAGLELLSRYED 57
             L TL E L S ++      LK      D+ R  +++S +         K +A     L+  E
Sbjct:   434 LATLEEAL-SEKERIIERLKEQRERDDRERLEEIESFRKENKDLKEKVNALQAELTEKES 492

Query:    58 TWAALHRRAKDCASAGELVDSEVVMLSAHWEKKKTSLVELQEQLQQLPALIADLESMTAN 117
             +   L   A   ASAG    DS++  L     E+KK    +L+ QL++     I D    M
Sbjct:   493 SLIDLKEHASSLASAGLKRDSKLKSLEIAIEQKKEECSKLEAQLKKAHN-IEDDSRMNPE 551

Query:   118 LTHLEASFEEVENNLLHLEDLCG--QCELERCKHMQSQQLENYKKNKRK---ELETFKAE 172
                     ++++     +   D CG   Q E++R     +   +++EN K +K K    ELE+
Sbjct:   552 FAD---QIKQLDKEASYYRDECGKAQAEVDRLLEIL-KEVENEKNDKDKKIAELESLTLR 607

Query:   173 LDAEHAQKVLEMEHTQQMKLKERQKFFEEAFQQDMEQYLSTGYLQIAE 220
                  +   +KV   ++H  QQ++  K+   +    EE  +++        ++ +LQI E
Sbjct:   608 HMKDQNKKVANLKHNQQLEKKKNAQLLEEVRRREDSMADNSQHLQIEE 655
```

Score = 100 (15.0 bits), Expect = 6.2e-02, P = 6.0e-02
Identities = 44/156 (28%), Positives = 76/156 (48%)

```
Query:    57 DTWAALHRRAKDCASAGELVDSEVVMLSAHWEKKKTSLVELQEQLQQLPAL-IADLESMT 115
             D  A+ +R     +C  A    VD  + +L     E +K      +     +L+ L     + D
Sbjct:   560 DKEASYYR--DECGKAQAEVDRLLEILK-EVENEKNDKDKKIAELESLTLRHMKDQNKKV 616

Query:   116 ANLTHLEASFEEVENNLLHLEDLCGQCE--LERCKHMQSQQLENYKKNKRKELETFKAEL 173
```

```
                ANL H +    E+ +N  L LE++  + +   +  +H+Q ++L N  +  R+EL+  KA L
Sbjct:      617 ANLKHNQ-QLEKKKNAQL-LEEVRRREDSMADNSQHLQIEELMNALEKTRQELDATKARL 674

Query:      174 DAEHAQKVLEME-HTQQMKLKERQKFFEEAFQQDMEQYLS 212
                 A   Q + E E H   +++ ER+K  EE  +   E  L+
Sbjct:      675 -ASTQQSLAEKEAHLANLRI-ERRKQLEEILEMKQEALLA 712
```

Pedant information for DKFZphfbr2_2k19, frame 2

Report for DKFZphfbr2_2k19.2

[LENGTH]	303	
[MW]	34814.78	
[pI]	5.23	
[PROSITE]	LEUCINE_ZIPPER 1	
[KW]	All_Alpha	
[KW]	LOW_COMPLEXITY	3.63 %
[KW]	COILED_COIL	14.52 %

```
SEQ    MLETLRERLLSVQQDFTSGLKTLSDKSREAKVKSKPRTVPFLPKYSAGLELLSRYEDTWA
SEG    ............................................................
PRD    ccchhhhhhhhhhhcccccchhhhhhhhhhhcccccccccchhhhhhhhhhhhchhh
COILS  ............................................................

SEQ    ALHRRAKDCASAGELVDSEVVMLSAHWEKKKTSLVELQEQLQQLPALIADLESMTANLTH
SEG    ................................xxxxxxxxxxx.................
PRD    hhhhhhhhchhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  ................................CCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ    LEASFEEVENNLLHLEDLCGQCELERCKHMQSQQLENYKKNKRKELETFKAELDAEHAQK
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  CCCCCCCCCCCCCCCCCC..........................................

SEQ    VLEMEHTQQMKLKERQKFFEEAFQQDMEQYLSTGYLQIAERREPIGSMSSMEVNVDMLEQ
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhccccccccccchhhhhhhh
COILS  ............................................................

SEQ    MVLMDISDQEALDVFLNSGGEENTVLSPALGRVDKLALAEPGQYRCHSPPKVRRENHLPV
SEG    ............................................................
PRD    hhhhhhchhhhhhhhhccccccceeeccccccccceeeccccccccccccccceeecccccc
COILS  ............................................................

SEQ    TYA
SEG    ...
PRD    ccc
COILS  ...
```

Prosite for DKFZphfbr2_2k19.2

PS00029	97->119	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_2k19.2)

DKFZphfbr2_2k14
----

group: cell cycle

DKFZphfbr2_2k14 encodes a novel 335 amino acid protein with strong similarity to rattus rattus IAG2 "implantation-associated protein" and the human N33 tumour-suppressor gene.

Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. In addition, the novel protein contains a RGD cell attachment site. Therefore the novel protein is a new putative tumour-suppressor gene.

The new protein can find application in modulating/blocking the cell cycle and in the therapy of tumours.

strong similarity to human N33 tumor suppressor gene

complete cDNA, complete cds, EST hits,
potential start at Bp 30 matches kozak consensus ANCatgG
potential transmembran protein (4 TM)
similarity to yeast OST3p (oligosaccharyltransferase gamma chain)

Sequenced by Qiagen

Locus: unknown

Insert length: 2241 bp
Poly A stretch at pos. 2221, no polyadenylation signal found

```
   1 TGGGACTTAT AGAAGGGAGA GGAGCGAACA TGGCAGCGCG TTGGCGGTTT
  51 TGGTGTGTCT CTGTGACCAT GGTGGTGGCG CTGCTCATCG TTTGCGACGT
 101 TCCCTCAGCC TCTGCCCAAA GAAAGAAGGA GATGGTGTTA TCAGAAAAGG
 151 TTAGTCAGCT GATGGAATGG ACTAACAAAA GACCTGTAAT AAGAATGAAT
 201 GGAGACAAGT TCCGTCGCCT TGTGAAAGCC CCACCGAGAA ATTACTCCGT
 251 TATCGTCATG TTCACTGCTC TCCAACTGCA TAGACAGTGT GTCGTTTGCA
 301 AGCAAGCTGA TGAAGAATTC CAGATCCTGG CAAACTCCTG GCGATACTCC
 351 AGTGCATTCA CCAACAGGAT ATTTTTTGCC ATGGTGGATT TTGATGAAGG
 401 CTCTGATGTA TTTCAGATGC TAAACATGAA TTCAGCTCCA ACTTTCATCA
 451 ACTTTCCTGC AAAAGGGAAA CCCAAACGGG GTGATACATA TGAGTTACAG
 501 GTGCGGGGTT TTTCAGCTGA GCAGATTGCC CGGTGGATCG CCGACAGAAC
 551 TGATGTCAAT ATTAGAGTGA TTAGACCCCC AAATTATGCT GGTCCCCTTA
 601 TGTTGGGATT GCTTTTGGCT GTTATTGGTG GACTTGTGTA TCTTCGAAGA
 651 AGTAATATGG AATTTCTCTT TAATAAAACT GGATGGGCTT TTGCAGCTTT
 701 GTGTTTTGTG CTTGCTATGA CATCTGGTCA AATGTGGAAC CATATAAGAG
 751 GACCACCATA TGCCCATAAG AATCCCCACA CGGGACATGT GAATTATATC
 801 CATGGAAGCA GTCAAGCCCA GTTTGTAGCT GAAACACACA TTGTTCTTCT
 851 GTTTAATGGT GGAGTTACCT TAGGAATGGT GCTTTTGTGT GAAGCTGCTA
 901 CCTCTGACAT GGATATTGGA AAGCGAAAGA TAATGTGTGT GGCTGGTATT
 951 GGACTTGTTG TATTATTCTT CAGTTGGATG CTCTCTATTT TTAGATCTAA
1001 ATATCATGGC TACCCATACA GCTTTCTGAT GAGTTAAAAA GGTCCCAGAG
1051 ATATATAGAC ACTGGAGTAC TGGAAATTGA AAAACGAAAA TCGTGTGTGT
1101 TTGAAAAGAA GAATGCAACT TGTATATTCT GTATTACCTC TTTTTTTCAA
1151 GTGATTTAAA TAGTTAATCA TTTAACCAAA GAAGATGTGT AGTGCCTTAA
1201 CAAGCAATCC TCTGTCAAAA TCTGAGGTAT TTGAAAATAA TTATCCTCTT
1251 AACCTTCTCT TCCCAGTGAA CTTTATGGAA CATTTAATTT AGTACAATTA
1301 AGTATATTAT AAAAATTGTA AAACTACTAC TTTGTTTTAG TTAGAACAAA
1351 GCTCAAAACT ACTTTAGTTA ACTTGGTCAT CTGATCTTAT ATTGCCTTAT
1401 CCAAAGATGG GGAAAGTAAG TCCTGACCAG GTGTTCCCAC ATATGCCTGT
1451 TACAGATAAC TACATTAGGA ATTCATTCTT AGCTTCTTCA TCTTTGTGTG
1501 GATGTGTATA CTTTACGCAT CTTTCCTTTT GAGTAGAGAA ATTATGTGTG
1551 TCATGTGGTC TTCTGAAAAT GGAACACCAT TCTTCAGAGC ACACGTCTAG
1601 CCCTCAGCAA GACAGTTGTT TCTCCTCCTC CTTGCATATT TCCTACTGCG
1651 CTCCAGCCTG AGTGATAGAG TGAGACTCTG TCTCAAAAAA AAAGTATCTC
1701 TAAATACAGG ATTATAATTT CTGCTTGAGT ATGGTGTTAA CTACCTTGTA
1751 TTTAGAAAGA TTTCAGATTC ATTCCATCTC CTTAGTTTTC TTTTAAGGTG
1801 ACCCATCTGT GATAAAAATA TAGCTTAGTG CTAAAATCAG TGTAACTTAT
1851 ACATGGCCTA AAATGTTTCT ACAAATTAGA GTTTGTCACT TATTCCATTT
1901 GTACCTAAGA GAAAAATAGG CTCAGTTAGA AAAGGACTCC CTGGCCAGGC
1951 GCAGTGACTT ACGCCTGTAA TCTCAGCACT TTGGGAGGCC AAGGCAGGCA
2001 GATCACGAGG TCAGGAGTTC GAGACCATCC TGGCCAACAT GGTGAAACCC
2051 CGTCTCTACT AAAAATATAA AAATTAGCTG GGTGTGGTGG CAGGAGCCTG
2101 TAATCCCAGC TGCACAGGAG GCTGAGGCAC GAGAATCACT TGAACTCAGG
2151 AGATGGAGGT TTCAGTGAGC CGAGATCACG CCACTGCACT CCAGCCTGGC
2201 AACAGAGCGA GACTCCATCT CAAAAAAAAA AAAAAAAAAA A
```

BLAST Results

No BLAST result

Medline entries

96299740:
Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22.

97243398:
Tumour-suppressor genes in prostatic oncogenesis: a positional approach.

98334474:
Concordant methylation of the ER and N33 genes in glioblastoma multiforme.

Peptide information for frame 3

ORF from 30 bp to 1034 bp; peptide length: 335
Category: strong similarity to known protein

```
  1 MAARWRFWCV SVTMVVALLI VCDVPSASAQ RKKEMVLSEK VSQLMEWTNK
 51 RPVIRMNGDK FRRLVKAPPR NYSVIVMFTA LQLHRQCVVC KQADEEFQIL
101 ANSWRYSSAF TNRIFFAMVD FDEGSDVFQM LNMNSAPTFI NFPAKGKPKR
151 GDTYELQVRG FSAEQIARWI ADRTDVNIRV IRPPNYAGPL MLGLLLAVIG
201 GLVYLRRSNM EFLFNKTGWA FAALCFVLAM TSGQMWNHIR GPPYAHKNPH
251 TGHVNYIHGS SQAQFVAETH IVLLFNGGVT LGMVLLCEAA TSDMDIGKRK
301 IMCVAGIGLV VLFFSWMLSI FRSKYHGYPY SFLMS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2k14, frame 3

TREMBL:RNAF8554_1 gene: "IAG2"; product: "implantation-associated protein"; Rattus norvegicus implantation-associated protein (IAG2) mRNA, partial cds., N = 1, Score = 1560, P = 3.4e-160

PIR:G02297 gene N33 protein - human, N = 1, Score = 1256, P = 5.6e-128

TREMBL:HSN33S11_1 gene: "N33"; product: "N33 protein form 2"; Human N33 protein form 2 (N33) gene, exon 11 and complete cds., N = 1, Score = 1252, P = 1.5e-127

```
>TREMBL:RNAF8554_1 gene: "IAG2"; product: "implantation-associated protein";
     Rattus norvegicus implantation-associated protein (IAG2) mRNA, partial cds.
            Length = 308

  HSPs:

 Score = 1560 (234.1 bits), Expect = 3.4e-160, P = 3.4e-160
 Identities = 295/307 (96%), Positives = 299/307 (97%)

Query:    29 AQRKKEMVLSEKVSQLMEWTNKRPVIRMNGDKFRRLVKAPPRNYSVIVMFTALQLHRQCV 88
             AQRKKE VL EKV QLMEWTN+RPVIRMNGDKFR LVKAPPRNYSVIVMFTALQLHRQCV
Sbjct:     2 AQRKKEKVLVEKVIQLMEWTNQRPVIRMNGDKFRPLVKAPPRNYSVIVMFTALQLHRQCV 61

Query:    89 VCKQADEEFQILANSWRYSSAFTNRIFFAMVDFDEGSDVFQMLNMNSAPTFINFPAKGKP 148
             VCKQADEEFQILAN WRYSSAFTNRIFFAMVDFDEGSDVFQMLNMNSAPTFINFP KGKP
Sbjct:    62 VCKQADEEFQILANFWRYSSAFTNRIFFAMVDFDEGSDVFQMLNMNSAPTFINFPPKGKP 121

Query:   149 KRGDTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLLAVIGGLVYLRRS 208
             KR DTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLLAVIGGLVYLRRS
Sbjct:   122 KRADTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLLAVIGGLVYLRRS 181

Query:   209 NMEFLFNKTGWAFAALCFVLAMTSGQMWNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAE 268
             NMEFLFNKTGWAFAALCFVLAMTSGQMWNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAE
```

```
Sbjct:   182 NMEFLFNKTGWAFAALCFVLAMTSGQMWNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAE 241

Query:   269 THIVLLFNGGVTLGMVLLCEAATSDMDIGKRKIMCVAGIGLVVLFFSWMLSIFRSKYHGY 328
             THIVLLFNGGVTLGMVLLCEAA SDMDIGKR++MC+AGIGLVVLFFSWMLSIFRSKYHGY
Sbjct:   242 THIVLLFNGGVTLGMVLLCEAAASDMDIGKRRMMCIAGIGLVVLFFSWMLSIFRSKYHGY 301

Query:   329 PYSFLMS 335
             PYSFLMS
Sbjct:   302 PYSFLMS 308
```

Pedant information for DKFZphfbr2_2k14, frame 3

Report for DKFZphfbr2_2k14.3

```
[LENGTH]       335
[MW]           38036.83
[pI]           9.68
[HOMOL]        TREMBL:RNAF8554_1 gene: "IAG2"; product: "implantation-associated protein";
Rattus norvegicus implantation-associated protein (IAG2) mRNA, partial cds. 1e-161
[FUNCAT]       30.07 organization of endoplasmatic reticulum     [S. cerevisiae, YOR085w]
4e-14
[FUNCAT]       06.07 protein modification (glycolsylation, acylation, myristylation,
palmitylation, farnesylation and processing)      [S. cerevisiae, YOR085w] 4e-14
[FUNCAT]       01.05.01 carbohydrate utilization    [S. cerevisiae, YOR085w] 4e-14
[EC]           2.4.1.119 Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 1e-12

[PIRKW]        glycosyltransferase 1e-12
[PIRKW]        transmembrane protein 6e-69
[PIRKW]        hexosyltransferase 1e-12
[PROSITE]      RGD    1
[PROSITE]      MYRISTYL       4
[PROSITE]      AMIDATION      1
[PROSITE]      CK2_PHOSPHO_SITE       2
[PROSITE]      PKC_PHOSPHO_SITE       4
[PROSITE]      ASN_GLYCOSYLATION      2
[KW]           SIGNAL_PEPTIDE 30
[KW]           TRANSMEMBRANE  4
[KW]           LOW_COMPLEXITY      5.97 %

SEQ   MAARWRFWCVSVTMVVALLIVCDVPSASAQRKKEMVLSEKVSQLMEWTNKRPVIRMNGDK
SEG   ............................................................
PRD   ccccceeeeeehhhhhhhhhhhccccccchhhhhhhhhhhhhhhhhcccceeeeeecccc
MEM   ............................................................

SEQ   FRRLVKAPPRNYSVIVMFTALQLHRQCVVCKQADEEFQILANSWRYSSAFTNRIFFAMVD
SEG   ............................................................
PRD   ceeeeecccccccceeeehhhhhhcccceeeehhhhhhhhhhhhcccccccccceeeeeec
MEM   ............................................................

SEQ   FDEGSDVFQMLNMNSAPTFINFPAKGKPKRGDTYELQVRGFSAEQIARWIADRTDVNIRV
SEG   ............................................................
PRD   cccccceeeecccccccceeecccccccccccceeeeeeecchhhhhhhhhhhheeeee
MEM   ...........................................................M

SEQ   IRPPNYAGPLMLGLLLAVIGGLVYLRRSNMEFLFNKTGWAFAALCFVLAMTSGQMWNHIR
SEG   .....xxxxxxxxxxxxxxxxxxx....................................
PRD   eccccccchhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccceeec
MEM   MMMMMMMMMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMM...

SEQ   GPPYAHKNPHTGHVNYIHGSSQAQFVAETHIVLLFNGGVTLGMVLLCEAATSDMDIGKRK
SEG   ............................................................
PRD   cccccccccccccceeeeccchhhhhhhheeeeeeecchhhhhhhhhhhcccccccccc
MEM   ......................MMMMMMMMMMMMMMMMMMMMMMMMMMMM.........

SEQ   IMCVAGIGLVVLFFSWMLSIFRSKYHGYPYSFLMS
SEG   ...................................
PRD   eeeeccccceeeeeehhhhhhhhhhccccccccccc
MEM   MMMMMMMMMMMMMMMMMMMMMMMMMM.........
```

Prosite for DKFZphfbr2_2k14.3

PS00001	71->75	ASN_GLYCOSYLATION	PDOC00001
PS00001	215->219	ASN_GLYCOSYLATION	PDOC00001
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005

PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	111->114	PKC_PHOSPHO_SITE	PDOC00005
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00008	193->199	MYRISTYL	PDOC00008
PS00008	233->239	MYRISTYL	PDOC00008
PS00008	259->265	MYRISTYL	PDOC00008
PS00008	278->284	MYRISTYL	PDOC00008
PS00009	296->300	AMIDATION	PDOC00009
PS00016	150->153	RGD	PDOC00016

(No Pfam data available for DKFZphfbr2_2k14.3)

DKFZphfbr2_3c18

group: nucleic acid management

DKFZphfbr2_3c18 encodes a novel 448 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicase and RNA-dependent ATPase
from the DEAD box family
group helicases
Summary DKFZphfbr2_3c18 encodes a novel 448 amino acid protein with
similarity to DEAD-box subfamily ATP-dependent RNA helicases.
Deletion of the yeast homolouge DBP5 is lethal.

strong similarity to RNA helicase and RNA-dependent ATPase from the
DEAD box family

complete cDNA, EST hits
complete cds ATG at Bp 109

Sequenced by AGOWA

Locus: /map="87.50 cR from top of Chr16 linkage group"

Insert length: 1713 bp
Poly A stretch at pos. 1696, no polyadenylation signal found

```
   1 TGGGGTAGTG GGGCTGGAGC AGAGCCTGCC GCGAACCCCC GGAGCCCACG
  51 ATCCCTCGTG CCATCCCTCG AATCCACCAG CACGAGCGTC CCACCCGCGC
 101 CTGGGACCAT GGCCACTGAC TCATGGGCCC TGGCGGTGGA CGAGCAGGAA
 151 GCTGCGGCTG AGTCGTTGAG CAACTTGCAT CTTAAGGAAG AGAAAATCAA
 201 ACCAGATACC AATGGTGCTG TTGTCAAGAC CAATGCCAAT GCAGAGAAGA
 251 CAGATGAAGA AGAGAAAGAG GACAGAGCTG CCCAGTCCTT ACTCAACAAG
 301 CTGATCAGAA GCAACCTTGT TGATAACACA AACCAAGTGG AAGTCCTGCA
 351 GCGGGATCCA AACTCCCCTC TGTACTCGGT GAAGTCTTTT GAAGAGCTTC
 401 GGCTCCCACA GAACTTAATT GCCCAATCTC AGTCTGGTAC TGGTAAAACA
 451 GCTGCCTTCG TGCTGGCCAT GCTTAGCCAA GTAGAACCTG CAAACAAATA
 501 CCCCCAGTGT CTATGTCTCT CCCCAACGTA TGAGCTCGCC CTCCAAACAG
 551 GAAAAGTGAT TGAACAAATG GGCAAATTTT ACCCTGAACT GAAGCTAGCT
 601 TATGCTGTTC GAGGCAATAA ATTGGAAAGA GGCCAGAAGA TCAGTGAGCA
 651 GATTGTCATT GGCACCCCTG GGACTGTGCT GGACTGGTGC TCCAAGCTCA
 701 AGTTCATTGA TCCCAAGAAA ATCAAGGTGT TTGTTCTGGA TGAGGCTGAT
 751 GTCATGATAG CCACTCAGGG CCACCAAGAT CAGAGCATCC GCATCCAGAG
 801 GATGCTGCCC AGGAACTGCC AGATGCTGCT TTTCTCCGCC ACCTTTGAAG
 851 ACTCTGTGTG GAAGTTTGCC CAGAAAGTGG TCCCAGACCC AAACGTTATC
 901 AAACTGAAGC GTGAGGAAGA GACCCTGGAC ACCATCAAGC AGTACTATGT
 951 CCTGTGCAGC AGCAGAGACG AGAAGTTCCA GGCCTTGTGT AACCTCTACG
1001 GGGCCATCAC CATTGCTCAA GCCATGATCT TCTGCCATAC TCGCAAAACA
1051 GCTAGTTGGC TGGCAGCAGA GCTCTCAAAA GAAGGCCACC AGGTGGCTCT
1101 GCTGAGTGGG GAGATGATGG TGGAACAGAG GGCTGCAGTG ATTGAGCGCT
1151 TCCGAGAGGG CAAAGAGAAG GTTTTGGTGA CCACCAACGT GTGTGCCCGC
1201 GGCATTGATG TTGAACAAGT GTCTGTCGTC ATCAACTTTG ATCTTCCCGT
1251 GGACAAGGAC GGGAATCCTG ACAATGAGAC CTACCTGCAC CGGATCGGGC
1301 GCACGGGCCG CTTTGGCAAG AGGGGCCTGG CAGTGAACAT GGTGGACAGC
1351 AAGCACAGCA TGAACATCCT GAACAGAATC CAGGAGCATT TTAATAAGAA
1401 GATAGAAAGA TTGGACACAG ATGATTTGGA CGAGATTGAG AAAATAGCCA
1451 ACTGAGAAGC TCCACCAGCC ACTGATGCCA GCCCTGGCAC TGCCCCTGCA
1501 CAGGAGACAA GTGCGTTCAG GGCACAGGCC CCGACATCAC CCCAAGGACA
1551 ACGGCACAAG TAGAGAGAAA CTACCTACCT CACTTCAAAT TATGTTTGGA
1601 CTTGACAAAA ATGTATGCAA ATGATGGGGG ATGGTAGAAA AAAATTATTT
1651 ACACAACCTT GGAAGATTAG GCATGAATAC ACAGAGATTT ACCTTTAAAA
1701 AAAAAAAAAA AAA
```

BLAST Results

Entry G36496 from database EMBL:
SHGC-53094 Human Homo sapiens STS cDNA.
Length = 459
Minus Strand HSPs:
Score = 1693 (254.0 bits), Expect = 2.8e-70, P = 2.8e-70
Identities = 369/387 (95%), Positives = 369/387 (95%)

Entry G44014 from database EMBLNEW:
WIAF-3643-STS Human THudson SANGER Homo sapiens STS genomic, sequence tagged site.
Score = 901, P = 2.3e-35, identities = 183/185

Medline entries

94192995:
Gene 1994 Mar 25;140(2):171-177
Mouse erythroid cells express multiple putative RNA helicase genes exhibiting
high sequence conservation from yeast to mammals.

Peptide information for frame 1

ORF from 109 bp to 1452 bp; peptide length: 448
Category: strong similarity to known protein

```
  1 MATDSWALAV DEQEAAAESL SNLHLKEEKI KPDTNGAVVK TNANAEKTDE
 51 EEKEDRAAQS LLNKLIRSNL VDNTNQVEVL QRDPNSPLYS VKSFEELRLP
101 QNLIAQSQSG TGKTAAFVLA MLSQVEPANK YPQCLCLSPT YELALQTGKV
151 IEQMGKFYPE LKLAYAVRGN KLERGQKISE QIVIGTPGTV LDWCSKLKFI
201 DPKKIKVFVL DEADVMIATQ GHQDQSIRIQ RMLPRNCQML LFSATFEDSV
251 WKFAQKVVPD PNVIKLKREE ETLDTIKQYY VLCSSRDEKF QALCNLYGAI
301 TIAQAMIFCH TRKTASWLAA ELSKEGHQVA LLSGEMMVEQ RAAVIERFRE
351 GKEKVLVTTN VCARGIDVEQ VSVVINFDLP VDKDGNPDNE TYLHRIGRTG
401 RFGKRGLAVN MVDSKHSMNI LNRIQEHFNK KIERLDTDDL DEIEKIAN
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3c18, frame 1

PIR:I49731 RNA helicase - mouse, N = 2, Score = 1758, P = 3.8e-223

TREMBL:AF005239_1 gene: "Dbp80"; product: "DEAD-box helicase"; Drosophila melanogaster DEAD-box helicase (Dbp80) mRNA, complete cds., N = 2, Score = 1142, P = 1.8e-125

SWISSPROT:YB66_SCHPO PUTATIVE ATP-DEPENDENT RNA HELICASE C12C2.06., N = 2, Score = 911, P = 5.5e-103

PIR:S66920 probable RNA helicase CA5/6 - yeast (Saccharomyces cerevisiae), N = 2, Score = 887, P = 1.9e-98

>PIR:I49731 RNA helicase - mouse
Length = 478

HSPs:

Score = 1758 (263.8 bits), Expect = 3.8e-223, Sum P(2) = 3.8e-223
Identities = 338/349 (96%), Positives = 349/349 (100%)

```
Query:  100 PQNLIAQSQSGTGKTAAFVLAMLSQVEPANKYPQCLCLSPTYELALQTGKVIEQMGKFYP 159
            PQNLIAQSQSGTGKTAAFVLAMLS+VEPA++YPQCLCLSPTYELALQTGKVIEQMGKF+P
Sbjct:  130 PQNLIAQSQSGTGKTAAFVLAMLSRVEPADRYPQCLCLSPTYELALQTGKVIEQMGKFHP 189

Query:  160 ELKLAYAVRGNKLERGQKISEQIVIGTPGTVLDWCSKLKFIDPKKIKVFVLDEADVMIAT 219
            ELKLAYAVRGNKLERGQK+SEQIVIGTPGTVLDWCSKLKFIDPKKIKVFVLDEADVMIAT
Sbjct:  190 ELKLAYAVRGNKLERGQKVSEQIVIGTPGTVLDWCSKLKFIDPKKIKVFVLDEADVMIAT 249

Query:  220 QGHQDQSIRIQRMLPRNCQMLLFSATFEDSVWKFAQKVVPDPNVIKLKREEETLDTIKQY 279
```

```
                QGHQDQSIRIQR++PRNCQMLLFSATFEDSVWKFAQKVVPDPN+IKLKREEETLDTIKQY
Sbjct:     250  QGHQDQSIRIQRIVPRNCQMLLFSATFEDSVWKFAQKVVPDPNIIKLKREEETLDTIKQY 309

Query:     280  YVLCSSRDEKFQALCNLYGAITIAQAMIFCHTRKTASWLAAELSKEGHQVALLSGEMMVE 339
                YVLC++R+EKFQALCNLYGAITIAQAMIFCHTRKTASWLAAELSKEGHQVALLSGEMMVE
Sbjct:     310  YVLCNNREEKFQALCNLYGAITIAQAMIFCHTRKTASWLAAELSKEGHQVALLSGEMMVE 369

Query:     340  QRAAVIERFREGKEKVLVTTNVCARGIDVEQVSVVINFDLPVDKDGNPDNETYLHRIGRT 399
                QRAAVIERFREGKEKVLVTTNVCARGIDVEQVSVVINFDLPVDKDGNPDNETYLHRIGRT
Sbjct:     370  QRAAVIERFREGKEKVLVTTNVCARGIDVEQVSVVINFDLPVDKDGNPDNETYLHRIGRT 429

Query:     400  GRFGKRGLAVNMVDSKHSMNILNRIQEHFNKKIERLDTDDLDEIEKIAN 448
                GRFGKRGLAVNMVDSKHSMNILNRIQEHFNKKIERLDTDDLDEIEKIAN
Sbjct:     430  GRFGKRGLAVNMVDSKHSMNILNRIQEHFNKKIERLDTDDLDEIEKIAN 478

 Score = 419 (62.9 bits), Expect = 3.8e-223, Sum P(2) = 3.8e-223
 Identities = 94/136 (69%), Positives = 104/136 (76%)

Query:       1  MATDSWALAVDEQEAAAESLSNLHLKEEKIKPDTNGAVVKTNANAEKTDEEEKEDRAAQS 60
                MATDSWALAVDEQEAA +S+S+L +KEEK K DTNG V+KT+  AEKT+EEEKEDRAAQS
Sbjct:       1  MATDSWALAVDEQEAAVKSMSSLQIKEEKAKSDTNG-VIKTSTTAEKTEEEEKEDRAAQS 59

Query:      61  LLNKLIRSNLVDNTNQVEVLQRDPNSPLYSVKSFEELRL-PQNL---IAQSQSGTGKTAA 116
                LLNKLIRSNLVDNTNQVEVLQRDP+SPLYSVKSFEELRL PQ L    A   +   K
Sbjct:      60  LLNKLIRSNLVDNTNQVEVLQRDPSSPLYSVKSFEELRLKPQLLQGVYAMGFNRPSKIQE 119

Query:     117  FVLAMLSQVEPANKYPQ 133
                  L M+    P N  Q
Sbjct:     120  NALPMMLAEPPQNLIAQ 136
```

Pedant information for DKFZphfbr2_3c18, frame 1

Report for DKFZphfbr2_3c18.1

```
[LENGTH]       448
[MW]           50490.07
[pI]           5.83
[HOMOL]        PIR:I49731 RNA helicase - mouse 0.0
[FUNCAT]       98 classification not yet clear-cut   [S. cerevisiae, YOR046c] 1e-102
[FUNCAT]       04.01.04 rrna processing     [S. cerevisiae, YDR021w] 2e-65
[FUNCAT]       30.10 nuclear organization    [S. cerevisiae, YDR021w] 2e-65
[FUNCAT]       30.03 organization of cytoplasm     [S. cerevisiae, YJL138c] 1e-63
[FUNCAT]       05.04 translation (initiation, elongation and termination)  [S. cerevisiae,
YJL138c] 1e-63
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YDL160c] 2e-49
[FUNCAT]       j mrna translation and ribosome biogenesis   [H. influenzae, HI0231 RNA] 9e-48
[FUNCAT]       04.05.03 mrna processing (splicing)  [S. cerevisiae, YDL084w] 1e-43
[FUNCAT]       1 genome replication, transcription, recombination and repair     [H.
influenzae, HI0892] 3e-39
[FUNCAT]       06.10 assembly of protein complexes   [S. cerevisiae, YLL008w] 1e-35
[FUNCAT]       09.01 biogenesis of cell wall         [S. cerevisiae, YJL033w] 9e-27
[FUNCAT]       04.05.01.07 chromatin modification    [S. cerevisiae, YMR290c] 8e-26
[FUNCAT]       30.16 mitochondrial organization      [S. cerevisiae, YDR194c] 1e-23
[FUNCAT]       r general function prediction         [M. jannaschii, MJ1401] 9e-08
[FUNCAT]       11.10 cell death       [S. cerevisiae, YMR190c] 1e-05
[FUNCAT]       03.19 recombination and dna repair    [S. cerevisiae, YMR190c] 1e-05
[FUNCAT]       99 unclassified proteins      [S. cerevisiae, YIR002c] 7e-04
[BLOCKS]       BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW]        nucleus 4e-64
[PIRKW]        RNA binding 1e-64
[PIRKW]        DEAD box 4e-64
[PIRKW]        transmembrane protein 3e-22
[PIRKW]        DNA binding 2e-32
[PIRKW]        ATP 1e-101
[PIRKW]        purine nucleotide binding 4e-64
[PIRKW]        P-loop 1e-101
[PIRKW]        hydrolase 4e-43
[PIRKW]        protein biosynthesis 1e-64
[PIRKW]        ATP binding 2e-35
[SUPFAM]       WW repeat homology 3e-29
[SUPFAM]       translation initiation factor eIF-4A 1e-64
[SUPFAM]       DEAD/H box helicase homology 1e-101
[SUPFAM]       DNA helicase recG 2e-06
[SUPFAM]       unassigned DEAD/H box helicases 1e-101
[SUPFAM]       ATP-dependent RNA helicase DBP1 9e-33
```

```
[SUPFAM]        ATP-dependent RNA helicase DHH1 4e-48
[SUPFAM]        tobacco ATP-dependent RNA helicase DB10 3e-29
[PROSITE]       MYRISTYL        5
[PROSITE]       AMIDATION       1
[PROSITE]       CK2_PHOSPHO_SITE        6
[PROSITE]       GLYCOSAMINOGLYCAN       1
[PROSITE]       PKC_PHOSPHO_SITE        8
[PROSITE]       ASN_GLYCOSYLATION       1
[PFAM]          Helicases conserved C-terminal domain
[PFAM]          DEAD and DEAH box helicases
[KW]            Alpha_Beta
```

```
SEQ     MATDSWALAVDEQEAAAESLSNLHLKEEKIKPDTNGAVVKTNANAEKTDEEEKEDRAAQS
PRD     ccchhhhhhhhhhhhhhhhccccchhhhhhhcccccceeeeeehhhhhhhhhhhhhhhhh

SEQ     LLNKLIRSNLVDNTNQVEVLQRDPNSPLYSVKSFEELRLPQNLIAQSQSGTGKTAAFVLA
PRD     hhhhhhhhcccccceeeeeecccccceeehhhhhhhhccceeeeecccccchhhhhh

SEQ     MLSQVEPANKYPQCLCLSPTYELALQTGKVIEQMGKFYPELKLAYAVRGNKLERGQKISE
PRD     hhhhhhhhccceeeeecchhhhhhhhhhhhhhcccccccceeeccccchhhhhhhe

SEQ     QIVIGTPGTVLDWCSKLKFIDPKKIKVFVLDEADVMIATQGHQDQSIRIQRMLPRNCQML
PRD     eeeecccccchhhhhhhhhhccccceeeeeecchhhhhhccchhhhhhhhhhcccccceee

SEQ     LFSATFEDSVWKFAQKVVPDPNVIKLKREEETLDTIKQYYVLCSSRDEKFQALCNLYGAI
PRD     eeeccccchhhhhhhhhhccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhch

SEQ     TIAQAMIFCHTRKTASWLAAELSKEGHQVALLSGEMMVEQRAAVIERFREGKEKVLVTTN
PRD     hhhhhheeecchhhhhhhhhhhhhhcccceeeeeeccccchhhhhhhhhhhhcccccceeeeeec

SEQ     VCARGIDVEQVSVVINFDLPVDKDGNPDNETYLHRIGRTGRFGKRGLAVNMVDSKHSMNI
PRD     cccccccceeeeeeeeeccccccccccccccceeeeeeccccccccccceeeeeeeecccchhhh

SEQ     LNRIQEHFNKKIERLDTDDLDEIEKIAN
PRD     hhhhhhhhhhhcccccccccchhhhhccc
```

Prosite for DKFZphfbr2_3c18.1

PS00001	389->393	ASN_GLYCOSYLATION	PDOC00001
PS00002	109->113	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	90->93	PKC_PHOSPHO_SITE	PDOC00005
PS00005	111->114	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	226->229	PKC_PHOSPHO_SITE	PDOC00005
PS00005	275->278	PKC_PHOSPHO_SITE	PDOC00005
PS00005	284->287	PKC_PHOSPHO_SITE	PDOC00005
PS00005	311->314	PKC_PHOSPHO_SITE	PDOC00005
PS00005	399->402	PKC_PHOSPHO_SITE	PDOC00005
PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	93->97	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	189->193	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	284->288	CK2_PHOSPHO_SITE	PDOC00006
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	175->181	MYRISTYL	PDOC00008
PS00008	185->191	MYRISTYL	PDOC00008
PS00008	385->391	MYRISTYL	PDOC00008
PS00008	406->412	MYRISTYL	PDOC00008
PS00009	402->406	AMIDATION	PDOC00009

Pfam for DKFZphfbr2_3c18.1

```
HMM_NAME     DEAD and DEAH box helicases

HMM              *gLpPWILRnIyeMGFEkPTPIQQqAIPiILeG....RDVMACAQTGSGK
                  ++ ++ +N ++          P          E+     +++A++Q+G+GK
Query         65  LIRSNLVDNTNQVEVLQRDPNSPLYSVKSFEELRLPQNLIAQSQSGTGK    113

HMM              TAAFlIPMLQHIDwdPWpqpPQdPrALILAPTRELAMQIQEEcRkFgkHM
                 TAAF++ ML+++ +      + PQ   +L L+PT ELA+Q+ ++++++GK++
Query        114 TAAFVLAMLSQVEPAN--KYPQ---CLCLSPTYELALQTGKVIEQMGKFY    158

HMM              ngIRImcIYGGtnMRdQMRmLeRGpPHIVIATPGRLIDHIER.gtldLDr
                 + ++ + ++      ++ +++ +++  +IVI+TPG ++D   +   +D ++
```

```
Query         159 PELKLAYAVR----GNKLERGQKISEQIVIGTPGTVLDWCSKLKFIDPKK    204

HMM               IeMLVMDEADRMLD.MGFIDQIRrIMrqIPMpwNRQTMMFSATMPdeIqE
                  I+++V+DEAD M+  +G +DQ  RI R++P  +N Q ++FSAT+ D++ +
Query         205 IKVFVLDEADVMIATQGHQDQSIRIQRMLP--RNCQMLLFSATFEDSVWK    252

HMM               LARrFMRNPIRInIdMdElTtnEnIkQwYiyVerEMWKfdcLcrLIe*
                  +A ++   +P  I ++++E T++ +IKQ+Y+ + + ++KF +LC+L++
Query         253 FAQKVVPDPNVIKLKREEETLD-TIKQYYVLCSSRDEKFQALCNLYG    298


HMM_NAME      Helicases conserved C-terminal domain

HMM               *EileeWLknlGIrvmYIHGdMpQeERdeIMddFNnGEynVLIcTDVggR
                   +L+ +L+++G +V+ + G M+ E+R  ++++F++G+ +VL++T+V +R
Query         316  SWLAAELSKEGHQVALLSGEMMVEQRAAVIERFREGKEKVLVTTNVCAR    364

HMM               GIDIPdVNHVINYDM....PWNPEq..YIQRIGRTgRIG*
                  GID+++V++VIN+D+     + NP++  Y++RIGRTGR+G
Query         365 GIDVEQVSVVINFDLPVDKDGNPDNETYLHRIGRTGRFG    403
```

Medline
PMID: 10322435
"Unwinding RNA in : DEAD-box proteins and related families." de la Cruz J, Kressler D, Linder P

DKFZphfbr2_3f16
----

group: brain derived

DKFZphfbr2_3f16 encodes a novel 127 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 1514 bp
Poly A stretch at pos. 1454, polyadenylation signal at pos. 1434

```
   1 GGGGGGACTG GAGAAGGGAG GCGGCGGGCG AAGCGCACGT CGAGCGGGGG
  51 AGCGGCGCTG CCTGTGGAGA TCCGCGGAGG CCGACAGGAT TCGTTGGCTG
 101 CCGTCCCCGC TGCTGTGCAT TGGGTTAAAA ACGACAACCA ACATCAGCCA
 151 TGAAAGATCC AAGTCGCAGC AGTACTAGCC CAAGCATCAT CAATGAAGAT
 201 GTGATTATTA ACGGTCATTC TCATGAAGAT GACAATCCAT TTGCAGAGTA
 251 CATGTGGATG GAAAATGAAG AAGAATTCAA CAGACAAATA GAAGAGGAGT
 301 TATGGGAAGA AGAATTTATT GAACGCTGTT TCCAAGAAAT GCTGGAAGAG
 351 GAAGAAGAGC ATGAATGGTT TATTCCAGCT CGAGATCTCC CACAAACTAT
 401 GGACCAAATC CAAGACCAGT TTAATGACCT TGTTATCAGT GAAGGCTCTT
 451 CTCTGGAAGA TCTTGTGGTC AAGAGCAATC TGAATCCAAA TGCAAAGGAG
 501 TTTGTTCCTG GGGTGAAGTA CGGAAATATT TGAGTAGACG GGGCCCTCTT
 551 TTGGTGGATG TAGCACAATT TCCACACTGT GAAGGCAGTA TTAGAAGACT
 601 TAATTGTAAA AGCACTCTTG TCACTGTGTT ACACTTATGC ATTGCCAAAG
 651 TTTTTGTTAG TCTTGCATGC TTAATAAAAG TGCTGAGACT GTTACTAAGT
 701 AAAAAGCTGT CAAACATTTA CTGAAAATAG AATTGGCCCC ATGGCTTGAT
 751 GTGAAGACAG CAAGGAAAGA AGCACCAGTC AAGTTGTGAA CAAGCACCAA
 801 ATTAAAAGAC CTAAACCTTA CCAAATTGTC TTTTTTTGAG GCTAATCTAT
 851 CACTTGTTAA TGTCTAAACT TTAAAATCAG TACATTTAAT TTGAGTTCCA
 901 ACTGTTAAGC ATATTTCTCA GACTTAAATT TGATTATGTC CCCATCAAAA
 951 AGAATCTCCA TTTTCTGAAG GTCTGTTAGT TAATTTGAGA TAATTTGTTA
1001 AAGGCAAGTA TGTCATATTA CTGAGGCTAC AAGTTAGTCA GCAGATGAGT
1051 GCCAGTCCAG CCTTTTCCGG TATGTTATTG TTAGAAATAT TGAGTTCTAA
1101 TGTTACATCT GAGGAAGTAT GTAATTTGAG AATTGTAACT TCTAAGGGAT
1151 TCACTGCATC ATAGCTATGC CTGTATGGAG TCTAACATAT GACCAATACC
1201 AACCCATAAT CCAGCTGAAC AAAGATACTG TAACATTATG ATTTGAGTGG
1251 TGCTTTTCCT TGCTTTGTTA ACCATCACGA GAGTCTGCAG CACAACTTTT
1301 AACAAAGCTA GAACAGTTTT GGCTTCTTAA ACTTCATATT TGGGTAGGTT
1351 AAGCTGCCAT ACGTGTTCAG TGTGAATAGT GTTTAAGTTG AAAATATTGT
1401 AAAAAAATTA TATTTTTTCA AAAATATTTA AAAAAATAAA TAATAGTAGA
1451 ACTGAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGAAAAA
1501 AAAAAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 150 bp to 530 bp; peptide length: 127
Category: putative protein

```
   1 MKDPSRSSTS PSIINEDVII NGHSHEDDNP FAEYMWMENE EEFNRQIEEE
```

```
 51 LWEEEFIERC FQEMLEEEEE HEWFIPARDL PQTMDQIQDQ FNDLVISEGS
101 SLEDLVVKSN LNPNAKEFVP GVKYGNI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3f16, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_3f16, frame 3

Report for DKFZphfbr2_3f16.3

```
[LENGTH]     127
[MW]         14998.41
[pI]         4.04
[BLOCKS]     BL01269D
[PROSITE]    MYRISTYL         1
[PROSITE]    CK2_PHOSPHO_SITE       2
[KW]         Alpha_Beta
[KW]         LOW_COMPLEXITY    27.56 %

SEQ    MKDPSRSSTSPSIINEDVIINGHSHEDDNPFAEYMWMENEEEFNRQIEEELWEEEFIERC
SEG    ....................................xxxxxxxxxxxxxxxxxxxxxxxx
PRD    cccccccccccccccccceeeeccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ    FQEMLEEEEEHEWFIPARDLPQTMDQIQDQFNDLVISEGSSLEDLVVKSNLNPNAKEFVP
SEG    xxxxxxxxxxx.................................................
PRD    hhhhhhhhhhhhcccccccccchhhhhhhhhcceeeecccccceeeeecccccccccccc

SEQ    GVKYGNI
SEG    .......
PRD    ccccccc
```

Prosite for DKFZphfbr2_3f16.3

```
PS00006      24->28    CK2_PHOSPHO_SITE      PDOC00006
PS00006     100->104   CK2_PHOSPHO_SITE      PDOC00006
PS00008     121->127   MYRISTYL              PDOC00008
```

(No Pfam data available for DKFZphfbr2_3f16.3)

DKFZphfbr2_3g8
----

group: metabolism

DKFZphfbr2_3g8.1 encodes a novel 178 amino acid protein with similarity to yeast ARD1 protein.

In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1. NAT1 controls full repression of the silent mating type locus HML, sporulation and entry into G0. ARD1 is involved in the assembly of the NAT 1-complex. The new protein could be part of this or an other NAT complex.

The new protein can find application modulating NAT assembly and action and therefore be important in metabolism of drugs and environmental mutagens.

strong similarity to N-TERMINAL ACETYLTRANSFERASE COMPLEX ARD1 homolog

complete cDNA, complete cds? start at Bp 40, EST hits

Sequenced by AGOWA

Locus: /map="20"

Insert length: 1030 bp
Poly A stretch at pos. 1013, no polyadenylation signal found

```
   1 TGGGCTTGGC GAACGGTCTT CGGAAGCGGC GGCGGCGCGA TGACCACGCT
  51 ACGGGCCTTT ACCTGCGACG ACCTGTTCCG CTTCAACAAC ATTAACTTGG
 101 ATCCACTTAC AGAAACTTAT GGGATTCCTT TCTACCTACA ATACCTCGCC
 151 CACTGGCCAG AGTATTTCAT TGTTGCAGTG GCACCTGGTG GAGAATTAAT
 201 GGGTTATATT ATGGGTAAAG CAGAAGGCTC AGTAGCTAGG GAAGAATGGC
 251 ACGGGCACGT CACAGCTCTG TCTGTTGCCC CAGAATTTCG ACGCCTTGGT
 301 TTGGCTGCTA AACTTATGGA GTTACTAGAG GAGATTTCAG AAAGAAAGGG
 351 TGGGTTTTTT GTGGATCTCT TTGTAAGAGT ATCTAACCAA GTTGCAGTTA
 401 ACATGTACAA GCAGTTGGGC TACAGTGTAT ATAGGACGGT CATAGAGTAC
 451 TATTCGGCCA GCAACGGGGA GCCTGATGAG GACGCTTATG ATATGAGGAA
 501 AGCACTTTCC AGGGATACTG AGAAGAAATC CATCATACCA TTACCTCATC
 551 CTGTGAGGCC TGAAGACATT GAATAACCCT GGGCAGTGGT TCTTAGGCAG
 601 ATACTCTAGA TGCTTTATGG ACAATATTAT TTTCATTGGA TGATTCTGGA
 651 GCTCTATTAG GAGAAAAGTA ATCATTTTAG GTCTTAAAGA CTTCAAGAAA
 701 ATACAGGTTA TCAATTTATT TTAAATCTCA TTGTTTCCAG TTAGCAATAT
 751 CATACCTATT AAAGCTGTTC ATTGTAACAA AATTCAATCA AAAAGGCAGC
 801 TAGGTCAGAA GGAAACATAC CACTCTCATG GTTCATAGTA TTCACTGTAT
 851 GTATGCTAGG GAAAAGACTT GCTCCAGTCT CCTCCTCAGT TCTGTGCCTG
 901 AGAACCACTG CTGCATATAT TTGTTTTTAA ATTTTGTATT GAACTGTTAA
 951 TTGAAGCTTT AAAAGCATAT ATGAAATGTA TAAATCTAAG ATGTATAATA
1001 CATTATTGAC TCCAAAAAAA AAAAAAAAAA
```

BLAST Results

Entry HSG0101 from database EMBL:
human STS SHGC-35956.
Length = 401
Minus Strand HSPs:
Score = 1417 (212.6 bits), Expect = 9.3e-58, P = 9.3e-58
Identities = 301/311 (96%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 40 bp to 573 bp; peptide length: 178
Category: strong similarity to known protein

```
  1 MTTLRAFTCD DLFRFNNINL DPLTETYGIP FYLQYLAHWP EYFIVAVAPG
 51 GELMGYIMGK AEGSVAREEW HGHVTALSVA PEFRRLGLAA KLMELLEEIS
```

```
101 ERKGGFFVDL FVRVSNQVAV NMYKQLGYSV YRTVIEYYSA SNGEPDEDAY
151 DMRKALSRDT EKKSIIPLPH PVRPEDIE
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3g8, frame 1

TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4., N = 1, Score = 475, P = 3.2e-45

SWISSPROT:ARDH_LEIDO N-TERMINAL ACETYLTRANSFERASE COMPLEX ARD1 SUBUNIT HOMOLOG., N = 1, Score = 451, P = 1.1e-42

PIR:S69021 hypothetical protein YPR131c - yeast (Saccharomyces cerevisiae), N = 1, Score = 382, P = 2.3e-35

```
>TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal
     acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4.
            Length = 180

  HSPs:

 Score = 475 (71.3 bits), Expect = 3.2e-45, P = 3.2e-45
 Identities = 96/165 (58%), Positives = 118/165 (71%)

Query:     1 MTTLRAFTCDDLFRFNNINLDPLTETYGIPFYLQYLAHWPEYFIVAVAPGGE--LMGYIM 58
             MT  R F   DLF FNNINLDPLTET+ I FYL YL  WP   +V  +   +  LMGYIM
Sbjct:     1 MTDTRKFKATDLFSFNNINLDPLTETFNISFYLSYLNKWPSLCVVQESDLSDPTLMGYIM 60

Query:    59 GKAEGSVAREEWHGHVTALSVAPEFRRLGLAAKLMELLEEISERKGGFFVDLFVRVSNQV 118
             GK+EG+   +EWH HVTA++VAP  RRLGLA  +M+ LE +   +  FFVDLFVR SN +
Sbjct:    61 GKSEGT--GKEWHTHVTAITVAPNSRRLGLARTMMDYLETVGNSENAFFVDLFVRASNAL 118

Query:   119 AVNMYKQLGYSVYRTVIEYYSASNGEPDEDAYDMRKALSRDTEKKSI 165
             A++ YK LGYSVYR VI YYS  +G+ DED++DMRK LSRD  ++SI
Sbjct:   119 AIDFYKGLGYSVYRRVIGYYSNPHGK-DEDSFDMRKPLSRDVNRESI 164
```

Pedant information for DKFZphfbr2_3g8, frame 1

Report for DKFZphfbr2_3g8.1

```
[LENGTH]        178
[MW]            20338.24
[pI]            5.06
[HOMOL]         TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal
acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4. 7e-47
[FUNCAT]        06.07 protein modification (glycolsylation, acylation, myristylation,
palmitylation, farnesylation and processing)       [S. cerevisiae, YPR131c] 6e-37
[FUNCAT]        01.06.07 lipid, fatty-acid and sterol utilization    [S. cerevisiae, YHR013c]
4e-14
[FUNCAT]        30.03 organization of cytoplasm       [S. cerevisiae, YHR013c] 4e-14
[FUNCAT]        03.22 cell cycle control and mitosis  [S. cerevisiae, YHR013c] 4e-14
[FUNCAT]        r general function prediction         [M. jannaschii, MJ1530] 6e-09
[PIRKW]         acyltransferase 1e-12
[SUPFAM]        arrest-defective protein 1 1e-12
[SUPFAM]        Escherichia coli peptide N-acetyltransferase rimI 1e-07
[PROSITE]       CK2_PHOSPHO_SITE        3
[PROSITE]       PKC_PHOSPHO_SITE        3
[KW]            Alpha_Beta

SEQ    MTTLRAFTCDDLFRFNNINLDPLTETYGIPFYLQYLAHWPEYFIVAVAPGGELMGYIMGK
PRD    cccccccccccchhhhhhccccccccccccchhhhhhcccccceeeeeeccccceeeehhhh

SEQ    AEGSVAREEWHGHVTALSVAPEFRRLGLAAKLMELLEEISERKGGFFVDLFVRVSNQVAV
PRD    hcccccccccccceeeeehhhhhhhhcchhhhhhhhhhhhhhcccceeeeeeeeecchhhhh

SEQ    NMYKQLGYSVYRTVIEYYSASNGEPDEDAYDMRKALSRDTEKKSIIPLPHPVRPEDIE
PRD    hhhhhhccccchhhhhhccccccccccchhhhhhhhhhhhhhhhhcccccccccccccc
```

Prosite for DKFZphfbr2_3g8.1

PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	100->103	PKC_PHOSPHO_SITE	PDOC00005
PS00005	160->163	PKC_PHOSPHO_SITE	PDOC00005
PS00006	8->12	CK2_PHOSPHO_SITE	PDOC00006
PS00006	133->137	CK2_PHOSPHO_SITE	PDOC00006
PS00006	141->145	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphfbr2_3g8.1)

DKFZphfbr2_312
----

group: brain derived

DKFZphfbr2_312 encodes a novel 589 amino acid protein with weak similarity to S. cerevisiae ubiquitin-like protein DSK2.

Pfam predicts for this protein similarity to the ubiquitin family; No informative BLAST results; No predictive prosite or SCOP motive

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to ubiquitin-like protein DSK2 yeast

complete cDNA, complete cds, EST hits
Dsk2p is involved in spindel pole body SPB duplication, SPB = centomer
strong similarity to HRIHFB2157 human mRNA

Sequenced by AGOWA

Locus: unknown

Insert length: 2978 bp
Poly A stretch at pos. 2958, polyadenylation signal at pos. 2924

```
   1 GGGGGGAGGA AGCGGTGGCT GCTGCGGATG TCGGTGTGAG CGAGCGGCGC
  51 CTGAACACAC GGCGGCTGCC GAGCGCCTGA CCCGGGCCTG CGCCAGAGCC
 101 TGCACCGAGC TCCGGGGCCC CACACCCGCT ACGGTGGCCC TGCGCCCGTT
 151 GCTACTGAGG CGGCGTGCTC TGCATTCTTC GCTGTCCAGG CCTGCCGGCT
 201 CTGGTGTCTG CTGGCTCCTC CTTGCTCGCC TGCTCCCTCC TGCTTGCCTG
 251 AGTCACCGCC GCCGCCGCCG CCACAGCCAT GGCCGAGAGT GGTGAAAGCG
 301 GCGGTCCTCC GGGCTCCCAG GATAGCGCCG CCGGAGCCGA AGGTGCTGGC
 351 GCCCCCGCGG CCGCTGCCTC CGCGGAGCCC AAAATCATGA AAGTCACCGT
 401 GAAGACCCCG AAGGAAAAGG AGGAATTCGC CGTGCCCGAG AATAGCTCCG
 451 TCCAGCAGTT TAAGGAAGAA ATCTCTAAAC GTTTTAAATC ACATACTGAC
 501 CAACTTGTGT TGATATTTGC TGGAAAAATT TTGAAAGATC AAGATACCTT
 551 GAGTCAGCAT GGAATTCATG ATGGACTTAC TGTTCACCTT GTCATTAAAA
 601 CACAAAACAG GCCTCAGGAT CATTCAGCTC AGCAAACAAA TACAGCTGGA
 651 GGCAATGTTA CTACATCATC AACTCCTAAT AGTAACTCTA CATCTGGTTC
 701 TGCTACTAGC AACCCTTTTG GTTTAGGTGG CCTTGGGGGA CTTGCAGGTC
 751 TGAGTAGCTT GGGTTTGAAT ACTACCAACT TCTCTGAACT ACAGAGTCAG
 801 ATGCAGCGAC AACTTTTGTC TAACCCTGAA ATGATGGTCC AGATCATGGA
 851 AAATCCCTTT GTTCAGAGCA TGCTCTCAAA TCCTGACCTG ATGAGACAGT
 901 TAATTATGGC CAATCCACAA ATGCAGCAGT TGATACAGAG AAATCCAGAA
 951 ATTAGTCATA TGTTGAATAA TCCAGATATA ATGAGACAAA CGTTGGAACT
1001 TGCCAGGAAT CCAGCAATGA TGCAGGAGAT GATGAGGAAC CAGGACCGAG
1051 CTTTGAGCAA CCTAGAAAGC ATCCCAGGGG GATATAATGC TTTAAGGCGC
1101 ATGTACACAG ATATTCAGGA ACCAATGCTG AGTGCTGCAC AAGAGCAGTT
1151 TGGTGGTAAT CCATTTGCTT CCTTGGTGAG CAATACATCC TCTGGTGAAG
1201 GTAGTCAACC TTCCCGTACA GAAAATAGAG ATCCACTACC CAATCCATGG
1251 GCTCCACAGA CTTCCCAGAG TTCATCAGCT TCCAGCGGCA CTGCCAGCAC
1301 TGTGGGTGGC ACTACTGGTA GTACTGCCAG TGGCACTTCT GGGCAGAGTA
1351 CTACTGCGCC AAATTTGGTG CCTGGAGTAG GAGCTAGTAT GTTCAACACA
1401 CCAGGAATGC AGAGCTTGTT GCAACAAATA ACTGAAAACC CACAACTGAT
1451 GCAAAACATG TTGTCTGCCC CCTACATGAG AAGCATGATG CAGTCACTAA
1501 GCCAGAATCC TGACCTTGCT GCACAGATGA TGCTGAATAA TCCCCTATTT
1551 GCTGGAAATC CTCAGCTTCA AGAACAAATG AGACAACAGC TCCCAACTTT
1601 CCTCCAACAA ATGCAGAATC CTGATACACT ATCAGCAATG TCAAACCCTA
1651 GAGCAATGCA GGCCTTGTTA CAGATTCAGC AGGGTTTACA GACATTAGCA
1701 ACGGAAGCCC CGGGCCTCAT CCCAGGGTTT ACTCCTGGCT TGGGGGCATT
1751 AGGAAGCACT GGAGGCTCTT CGGGAACTAA TGGATCTAAC GCCACACCTA
1801 GTGAAAACAC AAGTCCCACA GCAGGAACCA CTGAACCTGG ACATCAGCAG
1851 TTTATTCAGC AGATGCTGCA GGCTCTTGCT GGAGTAAATC CTCAGCTACA
1901 GAATCCAGAA GTCAGATTTC AGCAACAACT GGAACAACTC AGTGCAATGG
1951 GATTTTTGAA CCGTGAAGCA AACTTGCAAG CTCTAATAGC AACAGGAGGT
2001 GATATCAATG CAGCTATTGA AAGGTTACTG GGCTCCCAGC CATCATAGCA
2051 GCATTTCTGT ATCTTGAAAA AATGTAATTT ATTTTTGATA ACGGCTCTTA
2101 AACTTTAAAA TACCTGCTTT ATTTCATTTT GACTCTTGGA ATTCTGTGCT
2151 GTTATAAACA AACCCAATAT GATGCATTTT AAGGTGGAGT ACAGTAAGAT
2201 GTGTGGGTTT TTCTGTATTT TTCTTTTCTG GAACAGTGGG AATTAAGGCT
2251 ACTGCATGCA TCACTTCTGC ATTTATTGTA ATTTTTTAAA AACATCACCT
2301 TTTATAGTTG GGTGACCAGA TTTTGTCCTG CATCTGTCCA GTTTATTTGC
2351 TTTTTAAACA TTAGCCTATG GTAGTAATTT ATGTAGAATA AAAGCATTAA
2401 AAAGAAGCAA ATCATTTGCA CTCTATAATT TGTGGTACAG TATTGCTTAT
2451 TGTGACTTTG GCATGCATTT TTGCAAACAA TGCTGTAAGA TTTATACTAC
2501 TGATAATTTT GTTTTATTTG TATACAATAT AGAGTATGCA CATTTGGGAC
```

```
2551 TGCATTTCTG GAAACATACT GCAATAGGCT CTCTGAGCAA AACACCTGTA
2601 ACTAAAAAAG TGAAGATAAG AAAATACTCT TAAAGCTGAG TATTTCCTAA
2651 TTGTATAGAA TCTTACAGCA TCTTTGACAA ACATCTCCCA GCAAAAGTGC
2701 CGGTTAGTCA GGTTTGTTGA AAATACAGTA GAAAAGCTGA TTCTGGTTAT
2751 CTCTTTAAGG ACAATTAATT GTACAGACAC ATAATGTAAC ATTGTCTCAA
2801 CATTCATTCA CAGATTGACT GTAAATTACC TTAATCTTTG TGCAGACTGA
2851 AGGAACACTG TAGTATACCC CAAAGTGCAT TTGCCTAGGA CTTCTCAGCT
2901 TCTCCCATAG GTAGTTTAAC AGGCATTAAA ATTTGTAATT GAAATGTTGC
2951 TTTCACTCAA AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 279 bp to 2045 bp; peptide length: 589
Category: similarity to known protein

```
  1 MAESGESGGP PGSQDSAAGA EGAGAPAAAA SAEPKIMKVT VKTPKEKEEF
 51 AVPENSSVQQ FKEEISKRFK SHTDQLVLIF AGKILKDQDT LSQHGIHDGL
101 TVHLVIKTQN RPQDHSAQQT NTAGGNVTTS STPNSNSTSG SATSNPFGLG
151 GLGGLAGLSS LGLNTTNFSE LQSQMQRQLL SNPEMMVQIM ENPFVQSMLS
201 NPDLMRQLIM ANPQMQQLIQ RNPEISHMLN NPDIMRQTLE LARNPAMMQE
251 MMRNQDRALS NLESIPGGYN ALRRMYTDIQ EPMLSAAQEQ FGGNPFASLV
301 SNTSSGEGSQ PSRTENRDPL PNPWAPQTSQ SSSASSGTAS TVGGTTGSTA
351 SGTSGQSTTA PNLVPGVGAS MFNTPGMQSL LQQITENPQL MQNMLSAPYM
401 RSMMQSLSQN PDLAAQMMLN NPLFAGNPQL QEQMRQQLPT FLQQMQNPDT
451 LSAMSNPRAM QALLQIQQGL QTLATEAPGL IPGFTPGLGA LGSTGGSSGT
501 NGSNATPSEN TSPTAGTTEP GHQQFIQQML QALAGVNPQL QNPEVRFQQQ
551 LEQLSAMGFL NREANLQALI ATGGDINAAI ERLLGSQPS
```

BLASTP hits

Entry CE1_1 from database TREMBL:
"F15C11.2"; Caenorhabditis elegans cosmid VF15C11L
Length = 293
Score = 454 (159.8 bits), Expect = 4.4e-43, P = 4.4e-43
Identities = 81/162 (50%), Positives = 113/162 (69%)

Entry S54583 from database PIR:
ubiquitin-like protein DSK2 - yeast (Saccharomyces cerevisiae)
Length = 373
Score = 278 (97.9 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 100/307 (32%), Positives = 155/307 (50%)

Entry AB015344_1 from database TREMBLNEW:
gene: "HRIHFB2157"; Homo sapiens HRIHFB2157 mRNA, partial cds.
Score = 1135, P = 3.6e-115, identities = 227/301, positives = 253/301

Alert BLASTP hits for DKFZphfbr2_312, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_312, frame 3

Report for DKFZphfbr2_312.3

[LENGTH]	589
[MW]	62489.22
[pI]	5.02
[HOMOL]	TREMBL:AB015344_1 gene: "HRIHFB2157"; Homo sapiens HRIHFB2157 mRNA, partial cds. 1e-121
[FUNCAT]	03.22 cell cycle control and mitosis [S. cerevisiae, YMR276w] 2e-17

```
[FUNCAT]     30.10 nuclear organization    [S. cerevisiae, YMR276w] 2e-17
[BLOCKS]     BL00299 Ubiquitin family proteins
[SUPFAM]     unassigned ubiquitin-related proteins 5e-16
[SUPFAM]     ubiquitin homology 5e-16
[PROSITE]    MYRISTYL          24
[PROSITE]    CK2_PHOSPHO_SITE       9
[PROSITE]    GLYCOSAMINOGLYCAN      1
[PROSITE]    PKC_PHOSPHO_SITE       3
[PROSITE]    ASN_GLYCOSYLATION      7
[PFAM]       Ubiquitin family
[KW]         Irregular
[KW]         3D
[KW]         LOW_COMPLEXITY    23.43 %
```

```
SEQ    MAESGESGGPPGSQDSAAGAEGAGAPAAAASAEPKIMKVTVKTPKEKEEFAVPENSSVQQ
SEG    ..xxxxxxxxxxx..xxxxxxxxxxxxxxxxxx...xxxxxxxxxxxx............
laarA  ....................................CEEEEEETTTCEEEECTTTTBHHH

SEQ    FKEEISKRFKSHTDQLVLIFAGKILKDQDTLSQHGIHDGLTVHLVIKTQNRPQDHSAQQT
SEG    ............................................................
laarA  HHHHHHHHCCCGGGEEEEETTEECTTTTBGGGGCCTTTTEEEEEBC..............

SEQ    NTAGGNVTTSSTPNSNSTSGSATSNPFGLGGLGGLAGLSSLGLNTTNFSELQSQMQRQLL
SEG    ...xxxxxxxxxxxxxxxxxxxxx..xxxxxxxxxxxxxxxx..................
laarA  ............................................................

SEQ    SNPEMMVQIMENPFVQSMLSNPDLMRQLIMANPQMQQLIQRNPEISHMLNNPDIMRQTLE
SEG    ............................................................
laarA  ............................................................

SEQ    LARNPAMMQEMMRNQDRALSNLESIPGGYNALRRMYTDIQEPMLSAAQEQFGGNPFASLV
SEG    ............................................................
laarA  ............................................................

SEQ    SNTSSGEGSQPSRTENRDPLPNPWAPQTSQSSSASSGTASTVGGTTGSTASGTSGQSTTA
SEG    ..........................xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
laarA  ............................................................

SEQ    PNLVPGVGASMFNTPGMQSLLQQITENPQLMQNMLSAPYMRSMMQSLSQNPDLAAQMMLN
SEG    ............................................................
laarA  ............................................................

SEQ    NPLFAGNPQLQEQMRQQLPTFLQQMQNPDTLSAMSNPRAMQALLQIQQGLQTLATEAPGL
SEG    ............................................................
laarA  ............................................................

SEQ    IPGFTPGLGALGSTGGSSGTNGSNATPSENTSPTAGTTEPGHQQFIQQMLQALAGVNPQL
SEG    ....xxxxxxxxxxxxxxxxxxxxxxxx................................
laarA  ............................................................

SEQ    QNPEVRFQQQLEQLSAMGFLNREANLQALIATGGDINAAIERLLGSQPS
SEG    .................................................
laarA  .................................................
```

Prosite for DKFZphfbr2_312.3

PS00001	55->59	ASN_GLYCOSYLATION	PDOC00001
PS00001	126->130	ASN_GLYCOSYLATION	PDOC00001
PS00001	136->140	ASN_GLYCOSYLATION	PDOC00001
PS00001	164->168	ASN_GLYCOSYLATION	PDOC00001
PS00001	167->171	ASN_GLYCOSYLATION	PDOC00001
PS00001	302->306	ASN_GLYCOSYLATION	PDOC00001
PS00001	501->505	ASN_GLYCOSYLATION	PDOC00001
PS00002	305->309	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	40->43	PKC_PHOSPHO_SITE	PDOC00005
PS00005	43->46	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00006	43->47	CK2_PHOSPHO_SITE	PDOC00006
PS00006	71->75	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	200->204	CK2_PHOSPHO_SITE	PDOC00006
PS00006	260->264	CK2_PHOSPHO_SITE	PDOC00006
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00006	312->316	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	572->576	CK2_PHOSPHO_SITE	PDOC00006
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	12->18	MYRISTYL	PDOC00008

PS00008	19->25	MYRISTYL	PDOC00008
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	95->101	MYRISTYL	PDOC00008
PS00008	124->130	MYRISTYL	PDOC00008
PS00008	140->146	MYRISTYL	PDOC00008
PS00008	150->156	MYRISTYL	PDOC00008
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	162->168	MYRISTYL	PDOC00008
PS00008	267->273	MYRISTYL	PDOC00008
PS00008	293->299	MYRISTYL	PDOC00008
PS00008	308->314	MYRISTYL	PDOC00008
PS00008	337->343	MYRISTYL	PDOC00008
PS00008	343->349	MYRISTYL	PDOC00008
PS00008	347->353	MYRISTYL	PDOC00008
PS00008	355->361	MYRISTYL	PDOC00008
PS00008	366->372	MYRISTYL	PDOC00008
PS00008	479->485	MYRISTYL	PDOC00008
PS00008	489->495	MYRISTYL	PDOC00008
PS00008	492->498	MYRISTYL	PDOC00008
PS00008	495->501	MYRISTYL	PDOC00008
PS00008	499->505	MYRISTYL	PDOC00008
PS00008	573->579	MYRISTYL	PDOC00008

Pfam for DKFZphfbr2_312.3

```
HMM_NAME     Ubiquitin family

HMM              *MQIFVKTLtGRTcTFEVepQEtVeqIKQHIeekEGIPPeQQRLIFaGRQ
                  M ++VKT    +   +F V+++   V Q+K+ I+         +Q +LIFAG+
Query        37   MKVTVKTPK-EKEEFAVPENSSVQQFKEEISKRFKSHTDQLVLIFAGKI    84

HMM              LEDeKTLsDYNIggeSTLHLVlR*
                 L D  TLS+++I   + T+HLV++
Query        85  LKDQDTLSQHGIHDGLTVHLVIK     107
```

DKFZphfbr2_62b11
----

group: signal transduction

DKFZphfbr2_62b11.encodes a novel 655 amino acid putative GTPase-activating protein, related to human chimaerins.

The rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to CHIMAERIN

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="4"

Insert length: 4593 bp
Poly A stretch at pos. 4571, polyadenylation signal at pos. 4553

```
   1 GGGGGAGTTT GAAGACAGAA AGGAAAGGGG AGAAACCTGC AGAGAGCATC
  51 AAAGGATGGG GGGTGCTATA AAAGAAGCAG GGGGGTCCTT TGAAAGAAAT
 101 CTATCATGCA CTGAAATGCT TTCTGGAGAA GGTGCCGTTA TTTTCCTCCC
 151 CTCTTGCTCA GATGAAAGGA GCCAGCAAGG ACAGTCCTGA AATATTCCTC
 201 AGGGGACTTT TTGTCATTGT TCCTCTTTCC TCTTGCACAG AGCTATTTGC
 251 TGACCTTTCC AGAGGAATCT CAGTCCAGCT GAGAAGACAG TTCTTAATAA
 301 AAACAAAAAA ATGCAAAAAC CAATTCCTGC TGTTTGAATG GGAATGGTAG
 351 CTTGCTTGCT GCAGTTCTTT TCCTGTGACA TTTTGGAATG TCTGCAGAAA
 401 CTTAAAAAAA AGAAAAAAAA AACCTTAAAA ACTCCCTGGA TTAGGCAAGA
 451 GAAAAGGAAG TTTTTTTTTG CTAAACAGGA GTAAATGAGA GGTGGTAACT
 501 TATCCCTAAG CCAGGACCTG GATGATCAAA ACCTTCAAAT TCTAGGGATC
 551 AGCACTTCAA AAATAACAAG TAAACAAGCA TGAGGAGTGG CTGTTGGGTT
 601 TCGCTCAGAG GCAGGTTTTA AAGGAAGCCA AAACCGGGTT CAGAACTTCA
 651 GGCCTGTACG ATGCCTGAAG ACCGGAATTC TGGGGGGTGC CCGGCTGGTG
 701 CCTTAGCCTC AACTCCTTTC ATCCCTAAAA CTACATACAG AAGAATCAAA
 751 CGGTGTTTTA GTTTTCGGAA AGGCATTTTT GGACAGAAAC TGGAGGATAC
 801 TGTTCGTTAT GAGAAGAGAT ATGGGAACCG TCTGGCTCCG ATGTTGGTGG
 851 AGCAGTGCGT GGACTTTATC CGACAAAGGG GGCTGAAAGA AGAGGGTCTC
 901 TTTCGACTGC CAGGCCAGGC TAATCTTGTT AAGGAGCTCC AAGATGCCTT
 951 TGACTGTGGG GAGAAGCCAT CATTTGACAG CAACACAGAT GTACACACGG
1001 TGGCATCACT TCTTAAGCTG TACCTCCGAG AACTTCCAGA ACCAGTTATT
1051 CCTTATGCGA AGTATGAAGA TTTTTTGTCA TGTGCCAAAC TGCTCAGCAA
1101 GGAAGAGGAA GCAGGTGTTA AGGAATTAGC AAAGCAGGTG AAGAGTTTGC
1151 CAGTGGTAAA TTACAACCTC CTCAAGTATA TTTGCAGATT CTTGGATGAA
1201 GTACAGTCCT ACTCGGGAGT TAACAAAATG AGTGTGCAGA ACTTGGCAAC
1251 GGTCTTTGGT CCTAATATCC TGCGCCCCAA AGTGGAAGAT CCTTTGACTA
1301 TCATGGAGGG CACTGTGGTG GTCCAGCAGT TGATGTCAGT GATGATTAGC
1351 AAACATGATT GCCTCTTTCC CAAAGATGCA GAACTACAAA GCAAGCCCCA
1401 AGATGGAGTG AGCAACAACA ATGAAATTCA GAAGAAAGCC ACCATGGGGC
1451 TGTTACAGAA CAAGGAGAAC AATAACACCA AGGACAGCCC TAGTAGGCAG
1501 TGCTCCTGGG ACAAGTCTGA GTCACCCCAG AGAAGCAGCA TGAACAATGG
1551 ATCCCCCACA GCTCTATCAG GCAGCAAAAC CAACAGCCCA AAGAACAGTG
1601 TTCACAAGCT AGATGTGTCT AGAAGCCCCC CTCTCATGGT CAAAAAGAAC
1651 CCAGCCTTTA ATAAGGGTAG TGGGATAGTT ACCAATGGGT CCTTCAGCAG
1701 CAGTAATGCA GAAGGTCTTG AGAAAACCCA AACCACCCCC AATGGGAGCC
1751 TACAGGCCAG AAGGAGCTCT TCACTGAAGG TATCTGGTAC CAAAATGGGC
1801 ACGCACAGTG TACAGAATGG AACGGTGCGC ATGGGCATTT TGAACAGCGA
1851 CACACTCGGG AACCCCACAA ATGTTCGAAA CATGAGCTGG CTGCCAAATG
1901 GCTATGTGAC CCTGAGGGAT AACAAGCAGA AAGAACAAGC TGGAGAGTTA
1951 GGCCAGCACA ACAGACTGTC CACCTATGAT AATGTCCATC AACAGTTCTC
2001 CATGATGAAC CTTGATGACA AGCAGAGCAT TGACAGTGCT ACCTGGTCCA
2051 CTTCCTCCTG TGAAATCTCC CTCCCTGAGA ACTCCAACTC CTGTCGCTCT
2101 TCTACCACCA CCTGCCCAGA GCAAGACTTT TTTGGGGGGA ACTTTGAGGA
2151 CCCTGTTTTG GATGGGCCCC CGCAGGACGA CCTTTCCCAC CCCAGGGACT
2201 ATGAAAGCAA AAGTGACCAC AGGAGTGTGG GAGGTCGAAG TAGTCGTGCC
2251 ACCAGTAGCA GTGACAACAG TGAGACATTT GTGGGCAACA GCAGCAGCAA
2301 CCACAGTGCA CTGCACAGTT TAGTTTCCAG CCTGAAACAG GAAATGACCA
2351 AACAGAAGAT AGAGTATGAG TCCAGGATAA AGAGCTTAGA ACAGCGAAAC
2401 TTGACTTTGG AAACAGAAAT GATGAGCCTC CATGATGAAC TGGATCAGGA
2451 GAGGAAAAAG TTCACAATGA TAGAAATAAA AATGCGAAAT GCCGAGCGAG
2501 CAAAAGAAGA TGCCGAGAAA AGAAATGACA TGCTACAGAA AGAAATGGAG
2551 CAGTTTTTTT CCACGTTTGG AGAACTGACA GTGGAACCCA GGAGAACCGA
```

```
2601 GAGAGGAAAC ACAATATGGA TTCAGTGAGC CTGCTTTCGC CTGCTGTCTC
2651 TGATGGCTCT GGCAAGGACT CCAGGGATTC TGGTGGGATA TGACTTAGAA
2701 CCAGGTGGCT GGTCACCTGG ATGTACAGAA GTCTAACTGG TGAAGGAATA
2751 TCATTTACAG ACATTAAACA TCCATATCTG CAATGTGTAC CAAAGTTATA
2801 TCATGCCCCA TAATGCTACT GTCAAGTGTT ACAACTGGAT ATGTGTATAT
2851 AGAGTAGTTT TTCAAAAGTA AACTAAAAAT GAGAAGCATA TTTCAAGAAT
2901 TATTTTATTG CAAGTCTTGT ATTTAAATGT TAAATCAATA TGTTGTTGCA
2951 ATTTAGCTTG CTTTCAAGCT TCACCCCTTG CACTTAACAT AAGCTATTTT
3001 TGGCATTGTG TTATCATCGG CTTATTTTAT AGATCAATAT TTTTATTTCC
3051 CTTTTTTGCT GAGGAAATGA AGATAAGCAA AAATATAAAT ATATATATAA
3101 ATATATGAGT TATTAAAACC AGAAGAATAC TTTGTGGCTG TGCTGTTTGT
3151 GCCAATAGAC TTTGTCATGA CCAAAAAGAG AAATGTAAAT AGTTTTATAA
3201 AATACAGTCG AATCACCAGG AACCTTTGAG CTGCTTTTAA AATTCTTCCC
3251 CTGGCACCAC TCAGTTTTGC TTTTGCGAGG CGATTTGACA TAGGAACTTT
3301 GAGACTCCAT GAGAAAGTCC CTTTCTGAGG CCCACTGTCT ACCTTGCCAG
3351 ATCCTCAGTG CGTATCGCCA ATGCAGGATG CTCCTTAGAA AAGAAAAAAT
3401 GGTAAAGGAT GGCATTTAAC GATTCAGGCT TTGAATTACT CTGTCCCTCT
3451 GGACCGAATC TCTTTAACTG CTGGATAGTT TTAGAGGAAT TCTCCTGCTA
3501 CTTAGGTACT GGGAAACAAT GCTTGCTAAA CCATGCCCAC GTGAGCACCT
3551 GTCTCCCACT CAAACCTCTC CCATCTCCCA ACAACTGCAC TTTAGAATAC
3601 CAGCAGTGAA ATGGTATTAC TGTTTCCCTC TGAGTGAAAC TGCTAGAGTA
3651 TATGTCACGT AGTGACATTT TTTTCTCACT CAGGCTATTG CCATCTGGGA
3701 TTCTCTCCCT ACTACAGCTG GCAAAGTTGG TTTGCAGCAA GAAGATAGTG
3751 GGAGGGGGCC AGGCTGCAGG AGAAGGAGAA AAGTTTAGAA GAAACAAACC
3801 ATTTTGCTTC TAATTTTGAC AGTATCACTT TCCTGTTAAA ACATACAATA
3851 ATTTTAAAAG GTGAATGCCT AAAGTTCCAA TTTTAGCAAA TATGGGAACC
3901 TCAGCAATGC TAATTTTCTA GAAAAACCCA GGGCTCTTTG GAGCTAGAGT
3951 TTTGGGAGAA CAGTTCTTCA CAATAAGGCA ATGGTTTTGA GAGGCCAGGC
4001 AAATAATCTT TCTCACCGTA GAACAAAAAG TTACAAAAGG CATAATCGGA
4051 AATAGAGACT ACATACTTGA GTTTATGGGG TTTGTGTTGT TTGAAGGTTC
4101 AATGCTTGCA TGTGTTTATT TATTTTCAAG AGGGAAAGTG GTCTGTACTG
4151 CTTTCATCCT TGCCACTGTC TTGCTTTTAT TTTTTACTCT CCCACTGAGC
4201 AAGCGTCTGT GGTCCTATGG TATCAACCAG TATCTTTATA GCAATAATTT
4251 CTTTAATTCC CTTTTCTCTC TCTTTCCAAT TATTTAACCA GTTACTTCCA
4301 CCTGGACATA CGATAGGAAA TTCAAACTCA AAATATGAAA ATTGATCTTA
4351 ATAACTCTCC CTTCATATCT TTTCACCTAT TTCCAGTCCT TATCATAGTT
4401 GATAAAAACC TCAGACTCAT CCAGAAAGCT ATATGATGCA CTAGTAAAAA
4451 AAACAAAGAT ATTTAAACTG CTTGGGTTCA AATGGTATAC AATTTGCCAG
4501 CTGTTACTGA ACCTTCTATG CATAACTTTT TTTTTCCTCT GTGCAATTGG
4551 AATAATAAAA ATACTACTCC CATAAAAAAA AAAAAAAAAA AAC
```

BLAST Results

Entry G38474 from database EMBLNEW:
SHGC-58303 Human Homo sapiens STS genomic, sequence tagged site.
Score = 2175, P = 1.2e-92, identities = 439/441

Medline entries

97476250:
Beta2-chimaerin is a high affinity receptor for the phorbol ester tumor promoters.

Peptide information for frame 1

ORF from 661 bp to 2625 bp; peptide length: 655
Category: similarity to known protein

```
  1 MPEDRNSGGC PAGALASTPF IPKTTYRRIK RCFSFRKGIF GQKLEDTVRY
 51 EKRYGNRLAP MLVEQCVDFI RQRGLKEEGL FRLPGQANLV KELQDAFDCG
101 EKPSFDSNTD VHTVASLLKL YLRELPEPVI PYAKYEDFLS CAKLLSKEEE
151 AGVKELAKQV KSLPVVNYNL LKYICRFLDE VQSYSGVNKM SVQNLATVFG
201 PNILRPKVED PLTIMEGTVV VQQLMSVMIS KHDCLFPKDA ELQSKPQDGV
251 SNNNEIQKKA TMGLLQNKEN NNTKDSPSRQ CSWDKSESPQ RSSMNNGSPT
301 ALSGSKTNSP KNSVHKLDVS RSPPLMVKKN PAFNKGSGIV TNGSFSSSNA
351 EGLEKTQTTP NGSLQARRSS SLKVSGTKMG THSVQNGTVR MGILNSDTLG
401 NPTNVRNMSW LPNGYVTLRD NKQKEQAGEL GQHNRLSTYD NVHQQFSMMN
451 LDDKQSIDSA TWSTSSCEIS LPENSNSCRS STTTCPEQDF FGGNFEDPVL
501 DGPPQDDLSH PRDYESKSDH RSVGGRSSRA TSSSDNSETF VGNSSSNHSA
551 LHSLVSSLKQ EMTKQKIEYE SRIKSLEQRN LTLETEMMSL HDELDQERKK
```

```
601 FTMIEIKMRN AERAKEDAEK RNDMLQKEME QFFSTFGELT VEPRRTERGN
651 TIWIQ
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_62b11, frame 1

SWISSPROT:Y053_HUMAN HYPOTHETICAL PROTEIN KIAA0053., N = 3, Score = 661, P = 2.4e-89

TREMBL:HSU90908_1 product: "unknown"; Human clones 23549 and 23762 mRNA, complete cds., N = 1, Score = 348, P = 1.1e-29

PIR:S29128 N-chimerin - rat, N = 1, Score = 286, P = 2.8e-24

PIR:S29956 beta-chimerin - rat, N = 1, Score = 279, P = 1.6e-23

TREMBL:AB014572_1 gene: "KIAA0672"; product: "KIAA0672 protein"; Homo sapiens mRNA for KIAA0672 protein, complete cds., N = 1, Score = 314, P = 1e-24

```
>SWISSPROT:Y053_HUMAN HYPOTHETICAL PROTEIN KIAA0053.
          Length = 638

  HSPs:

 Score = 661 (99.2 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
 Identities = 122/209 (58%), Positives = 160/209 (76%)

Query:    38 GIFGQKLEDTVRYEKRYGNRLAPMLVEQCVDFIRQRGLKEEGLFRLPGQANLVKELQDAF 97
             G+FGQ+L++TV YE+++G  L P+LVE+C +FI + G  EEG+FRLPGQ NLVK+L+DAF
Sbjct:   148 GVFGQRLDETVAYEQKFGPHLVPILVEKCAEFILEHGRNEEGIFRLPGQDNLVKQLRDAF 207

Query:    98 DCGEKPSFDSNTDVHTVASLLKLYLRELPEPVIPYAKYEDFLSCAKLLSKEEEAGVKELA 157
             D GE+PSFD +TDVHTVASLLKLYLR+LPEPV+P+++YE FL C +L + +E    +EL
Sbjct:   208 DAGERPSFDRDTDVHTVASLLKLYLRDLPEPVVPWSQYEGFLLCGQLTNADEAKAQQELM 267

Query:   158 KQVKSLPVVNYNLLKYICRFLDEVQSYSGVNKMSVQNLATVFGPNILRPKVEDPLTIMEG 217
             KQ+  LP  NY+LL YICRFL E+Q    VNKMSV NLATV G N++R KVEDP  IM G
Sbjct:   268 KQLSILPRDNYSLLSYICRFLHEIQLNCAVNKMSVDNLATVIGVNLIRSKVEDPAVIMRG 327

Query:   218 TVVVQQLMSVMISKHDCLFPKDAELQSKP 246
             T  +Q++M++MI  H+ LFPK  ++   P
Sbjct:   328 TPQIQRVMTMMIRDHEVLFPKSKDIPLSP 356

 Score = 210 (31.5 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
 Identities = 45/115 (39%), Positives = 73/115 (63%)

Query:   531 TSSSDNSETFVGNSSSNHSALHSL---VSSLKQEMTKQKIEYESRIKSLEQRNLTLETEM 587
             T +S NSET  G  +S    + SL   V  L++E+  QK  YE +IK+LE+ N  +  ++
Sbjct:   523 TLASPNSETGPGKKNSGEEEIDSLQRMVQELRKEIETQKQMYEEQIKNLEKENYDVWAKV 582

Query:   588 MSLHDELDQERKKFTMIEIKMRNAERAKEDAEKRNDMLQKEMEQFFSTFGELTVE 642
             + L++EL++E+KK     +EI +RN ER++ED EKRN  L++E++++F  +  E  E
Sbjct:   583 VRLNEELEKEKKKSAALEISLRNMERSREDVEKRNKALEEEVKEFVKSMKEPKTE 637

 Score = 70 (10.5 bits), Expect = 1.2e-74, Sum P(3) = 1.2e-74
 Identities = 28/121 (23%), Positives = 54/121 (44%)

Query:   528 SRATSSSDNSETFVGNSSSNHSALHSLVSSLKQE-MTKQKIEYESRIKSLEQRNL-TLET 585
             S+ TS+ DN  +  G+     SAL S      K + +    E     K+  +  + +L+
Sbjct:   489 SQRTSTYDNVPSLPGSPGEEASALSSQACDSKGDTLASPNSETGPGKKNSGEEEIDSLQR 548

Query:   586 EMMSLHDELDQERKKFTMIEIKMRNAERAKEDAEKRNDMLQKEMEQFFSTFGELTVEPRR 645
              +  L  E++ +++   M E +++N E+   D   +   L +E+E+        L +  R
Sbjct:   549 MVQELRKEIETQKQ---MYEEQIKNLEKENYDVWAKVVRLNEELEKEKKKSAALEISLRN 605

Query:   646 TER 648
              ER
Sbjct:   606 MER 608

 Score = 53 (8.0 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
 Identities = 31/111 (27%), Positives = 46/111 (41%)

Query:   344 SFSSSNAEGLEKTQTTPNGSLQARRSSSLKVSGTKMGTHSVQNG----TV--RMGILNSD 397
             SFSS  ++  + T  T      A     S KV   K G   +Q+     T+  R   L S
Sbjct:   388 SFSSMTSDS-DTTSPTGQQPSDAFPEDSSKVPREKPGDWKMQSRKRTQTLPNRKCFLTSA 446
```

```
Query:   398 TLG-NPTNV---RNMSWLPNGYVTLRDNKQKEQAGELGQ---HNRLSTYDNV 442
             G N + +    +N  W P+       +  ++  + +L Q     R STYDNV
Sbjct:   447 FQGANSSKMEIFKNEFWSPSSEAKAGEGHRRTMSQDLRQLSDSQRTSTYDNV 498

 Score = 53 (8.0 bits), Expect = 3.5e-14, Sum P(3) = 3.5e-14
 Identities = 32/125 (25%), Positives = 56/125 (44%)

Query:   242 LQSKPQDG---VSNNNEIQKKATMGLLQNKEN--NNTKD---SPSRQCSWDKSESPQRSS 293
             ++SK +D     +     +IQ+  TM ++++ E     +KD   SP  Q +  K      RSS
Sbjct:   314 IRSKVEDPAVIMRGTPQIQRVMTM-MIRDHEVLFPKSKDIPLSPPAQKNDPKKAPVARSS 372

Query:   294 MNNGSPTALSGSKTNSPKNSVHKLDVSRSPPLMVKKNPAFNKGSGIVTNGSFSSSNAEGL 353
             +    +   L  S+T+S   +       D +   P    + + AF + S  V              +
Sbjct:   373 VGWDATEDLRISRTDSFSSMTSDSDTTS--PTGQQPSDAFPEDSSKVPREKPGDWKMQSR 430

Query:   354 EKTQTTPN 361
             ++TQT PN
Sbjct:   431 KRTQTLPN 438
```

Pedant information for DKFZphfbr2_62b11, frame 1

Report for DKFZphfbr2_62b11.1

```
[LENGTH]       655
[MW]           73394.60
[pI]           8.13
[HOMOL]        SWISSPROT:Y053_HUMAN HYPOTHETICAL PROTEIN KIAA0053. 3e-71
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
        [S. cerevisiae, YPL115c] 1e-16
[FUNCAT]       09.04 biogenesis of cytoskeleton      [S. cerevisiae, YPL115c] 1e-16
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YPL115c]
1e-16
[FUNCAT]       10.02.09 regulation of g-protein activity    [S. cerevisiae, YPL115c] 1e-16
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YER155c] 2e-16
[FUNCAT]       30.03 organization of cytoplasm      [S. cerevisiae, YER155c] 2e-16
[FUNCAT]       10.99 other signal-transduction activities    [S. cerevisiae, YDR379w] 4e-16
[FUNCAT]       03.10 sporulation and germination    [S. cerevisiae, YDL240w] 3e-15
[FUNCAT]       06.10 assembly of protein complexes  [S. cerevisiae, YOR134w] 2e-13
[FUNCAT]       30.04 organization of cytoskeleton   [S. cerevisiae, YOR134w] 2e-13
[SCOP]         d1rgp__ 1.83.1.1.1 p50 RhoGAP domain [human (Homo sapiens) 2e-46
[SCOP]         d1pbwa_ 1.83.1.1.2 p85 alpha subunit RhoGAP domain [human (Hom 6e-37
[PIRKW]        phosphotransferase 3e-13
[PIRKW]        breakpoint cluster region 2e-20
[PIRKW]        transmembrane protein 7e-14
[PIRKW]        brain 2e-20
[PIRKW]        alternative splicing 2e-20
[PIRKW]        P-loop 9e-19
[PIRKW]        cytoskeleton 1e-08
[SUPFAM]       CDC24 homology 7e-21
[SUPFAM]       bcr protein 7e-21
[SUPFAM]       myosin motor domain homology 9e-19
[SUPFAM]       pleckstrin repeat homology 2e-15
[SUPFAM]       LIM metal-binding repeat homology 9e-15
[SUPFAM]       protein kinase C zinc-binding repeat homology 5e-24
[PROSITE]      MYRISTYL         16
[PROSITE]      CAMP_PHOSPHO_SITE       3
[PROSITE]      CK2_PHOSPHO_SITE        15
[PROSITE]      TYR_PHOSPHO_SITE        2
[PROSITE]      PKC_PHOSPHO_SITE        11
[PROSITE]      ASN_GLYCOSYLATION       8
[KW]           Irregular
[KW]           3D
[KW]           LOW_COMPLEXITY      6.87 %
[KW]           COILED_COIL        12.06 %

SEQ     MPEDRNSGGCPAGALASTPFIPKTTYRRIKRCFSFRKGIFGQKLEDTVRYEKRYGNRLAP
SEG     ............................................................
COILS   ............................................................
1rgp-   ...........................................................C

SEQ     MLVEQCVDFIRQRGLKEEGLFRLPGQANLVKELQDAFDCGEKPSFDSNTDVHTVASLLKL
SEG     ............................................................
COILS   ............................................................
1rgp-   HHHHHHHHHHHHHHTTTTTTTTTCCCHHHHHHHHHHHHHCCCCCGGGCCCCHHHHHHHHH

SEQ     YLRELPEPVIPYAKYEDFLSCAKLLSKEEEAGVKELAKQVKSLPVVNYNLLKYICRFLDE
SEG     ............................................................
```

```
COILS   ............................................................
1rgp-   HHHHTTTTTTTGGGHHHHHH---TTTTCGGGHHHHHHHHHHHHCCHHHHHHHHHHHHHHH

SEQ     VQSYSGVNKMSVQNLATVFGPNILRPKVEDPLTIMEGTVVVQQLMSVMISKHDCLFPKDA
SEG     ............................................................
COILS   ............................................................
1rgp-   HHHHHHHHCCCHHHHHHHHGGGCC....................................

SEQ     ELQSKPQDGVSNNNEIQKKATMGLLQNKENNNTKDSPSRQCSWDKSESPQRSSMNNGSPT
SEG     ............................................................
COILS   ............................................................
1rgp-   ............................................................

SEQ     ALSGSKTNSPKNSVHKLDVSRSPPLMVKKNPAFNKGSGIVTNGSFSSSNAEGLEKTQTTP
SEG     ............................................................
COILS   ............................................................
1rgp-   ............................................................

SEQ     NGSLQARRSSSLKVSGTKMGTHSVQNGTVRMGILNSDTLGNPTNVRNMSWLPNGYVTLRD
SEG     ............................................................
COILS   ............................................................
1rgp-   ............................................................

SEQ     NKQKEQAGELGQHNRLSTYDNVHQQFSMMNLDDKQSIDSATWSTSSCEISLPENSNSCRS
SEG     .....................................................xxxxxxx
COILS   ............................................................
1rgp-   ............................................................

SEQ     STTTCPEQDFFGGNFEDPVLDGPPQDDLSHPRDYESKSDHRSVGGRSSRATSSSDNSETF
SEG     xxxxx...................................xxxxxxxxxxxxxxxxx...
COILS   ............................................................
1rgp-   ............................................................

SEQ     VGNSSSNHSALHSLVSSLKQEMTKQKIEYESRIKSLEQRNLTLETEMMSLHDELDQERKK
SEG     ..xxxxxxxxxxxxxxxx..........................................
COILS   ................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
1rgp-   ............................................................

SEQ     FTMIEIKMRNAERAKEDAEKRNDMLQKEMEQFFSTFGELTVEPRRTERGNTIWIQ
SEG     .......................................................
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC....................
1rgp-   .......................................................
```

Prosite for DKFZphfbr2_62b11.1

PS00001	271->275	ASN_GLYCOSYLATION	PDOC00001
PS00001	342->346	ASN_GLYCOSYLATION	PDOC00001
PS00001	361->365	ASN_GLYCOSYLATION	PDOC00001
PS00001	386->390	ASN_GLYCOSYLATION	PDOC00001
PS00001	407->411	ASN_GLYCOSYLATION	PDOC00001
PS00001	543->547	ASN_GLYCOSYLATION	PDOC00001
PS00001	547->551	ASN_GLYCOSYLATION	PDOC00001
PS00001	580->584	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	367->371	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	599->603	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	34->37	PKC_PHOSPHO_SITE	PDOC00005
PS00005	47->50	PKC_PHOSPHO_SITE	PDOC00005
PS00005	309->312	PKC_PHOSPHO_SITE	PDOC00005
PS00005	371->374	PKC_PHOSPHO_SITE	PDOC00005
PS00005	388->391	PKC_PHOSPHO_SITE	PDOC00005
PS00005	417->420	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	527->530	PKC_PHOSPHO_SITE	PDOC00005
PS00005	557->560	PKC_PHOSPHO_SITE	PDOC00005
PS00005	646->649	PKC_PHOSPHO_SITE	PDOC00005
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	213->217	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	348->352	CK2_PHOSPHO_SITE	PDOC00006
PS00006	417->421	CK2_PHOSPHO_SITE	PDOC00006
PS00006	437->441	CK2_PHOSPHO_SITE	PDOC00006
PS00006	465->469	CK2_PHOSPHO_SITE	PDOC00006
PS00006	470->474	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	516->520	CK2_PHOSPHO_SITE	PDOC00006
PS00006	532->536	CK2_PHOSPHO_SITE	PDOC00006

PS00006	589->593	CK2_PHOSPHO_SITE	PDOC00006
PS00006	602->606	CK2_PHOSPHO_SITE	PDOC00006
PS00006	635->639	CK2_PHOSPHO_SITE	PDOC00006
PS00007	43->51	TYR_PHOSPHO_SITE	PDOC00007
PS00007	176->185	TYR_PHOSPHO_SITE	PDOC00007
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	13->19	MYRISTYL	PDOC00008
PS00008	249->255	MYRISTYL	PDOC00008
PS00008	263->269	MYRISTYL	PDOC00008
PS00008	297->303	MYRISTYL	PDOC00008
PS00008	304->310	MYRISTYL	PDOC00008
PS00008	338->344	MYRISTYL	PDOC00008
PS00008	343->349	MYRISTYL	PDOC00008
PS00008	352->358	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00008	376->382	MYRISTYL	PDOC00008
PS00008	392->398	MYRISTYL	PDOC00008
PS00008	400->406	MYRISTYL	PDOC00008
PS00008	524->530	MYRISTYL	PDOC00008
PS00008	542->548	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_62b11.1)

DKFZphfbr2_62f10
----

group: intracellular transport and trafficking

DKFZphfbr2_62f10 encodes a novel 320 amino acid protein with strong similarity to mammalian zinc transporter proteins.

The novel proteins is a membrane protein, which should be involved in the transport of Zinc across the cell membrane.

The Zn-T-transporters are membrane proteins that facilitates sequestration of zinc in endosomal vesicles. In the brain, ZnT-3 mRNA seems to be involved in the accumulation of zinc in synaptic vesicles. Zinc (Zn) is an essential element in normal development and metabolism. Recent studies show that in Alzheimer's disease, Zn functions as a double-edged sword, affording protection against Alzheimer's amyloid beta peptide (the major component of senile plaques) at low concentrations and enhancing toxicity at high concentrations by accelerated aggregation of the amyloid beta peptide.

The new protein can find application in modulation of Zinc transport in neuronal cells, thus providing means for a modulation of Alzheimer's amyloid beta peptide plaque formation.

strong similarity to zinc transporter proteins ;
membrane regions: 5
Summary DKFZphfbr2_62f10 encodes a novel 320 amino acid protein with similarity to zinc transporter protein.
The new protein can find clinical application in modulating Zn2+ uptake.

strong similarity to zinc transporter proteins

complete cDNA, complete cds, few EST hits

Sequenced by LMU

Locus: unknown

Insert length: 5422 bp
Poly A stretch at pos. 5397, polyadenylation signal at pos. 5381

```
   1 GTCTAACTTT GGAAATATCA CCCTCATGCT GTCTTCCCAG GATGTCTCTC
  51 TCCCTAAGTA AGGGATGTTA CTTCCTGGAG GGAATGCAGT GTTGGGAATC
 101 TGAAGACCCA GCTTTGAGCT GAATTTGCTT TGTGATACCT GGAGAGAAGA
 151 CGTGTTTTCT TGACAACAGC ACAGTACCTA GTGAGTTCAA CAACAACGAC
 201 AACAACAGCC GCAGCTCATC CTGGCCGTCA TGGAGTTTCT TGAAAGAGCG
 251 TATCTTGTGA ATGATAAAGC TGCCAAGATG TATGCTTTCA CACTAGAAAG
 301 AAGGAGCTGC AAATGAACAC TTCATAGCAA TGTGGAACTC CAACAGAAAC
 351 CGGTGAATAA AGATCAGTGT CCCAGAGAGA GACCAGAGGA GCTGGAGTCA
 401 GGAGGCATGT ACCACTGCCA CAGTGGCTCC AAGCCCACAG AAAAGGGGGC
 451 GAATGAGTAC GCCTATGCCA AGTGGAAACT CTGTTCTGCT TCAGCAATAT
 501 GCTTCATTTT CATGATTGCA GAGGTCGTGG GTGGGCACAT TGCTGGGAGT
 551 CTTGCTGTTG TCACAGATGC TGCCCACCTC TTAATTGACC TGACCAGTTT
 601 CCTGCTCAGT CTCTTCTCCC TGTGGTTGTC ATCGAAGCCT CCCTCTAAGC
 651 GGCTGACATT TGGATGGCAC CGAGCAGAGA TCCTTGGTGC CCTGCTCTCC
 701 ATCCTGTGCA TCTGGGTGGT GACTGGCGTG CTAGTGTACC TGGCATGTGA
 751 GCGCCTGCTG TATCCTGATT ACCAGATCCA GGCGACTGTG ATGATCATCG
 801 TTTCCAGCTG CGCAGTGGCG GCCAACATTG TACTAACTGT GGTTTTGCAC
 851 CAGAGATGCC TTGGCCACAA TCACAAGGAA GTACAAGCCA ATGCCAGCGT
 901 CAGAGCTGCT TTTGTGCATG CCCCTGGAGA TCTATTTCAG AGTATCAGTG
 951 TGCTAATTAG TGCACTTATT ATCTACTTTA AGCCAGAGTA TAAAATAGCC
1001 GACCCAATCT GCACATTCAT CTTTTCCATC CTGGTCTTGG CCAGCACCAT
1051 CACTATCTTA AAGGACTTCT CCATCTTACT CATGGAAGGT GTGCCAAAGA
1101 GCCTGAATTA CAGTGGTGTG AAAGAGCTTA TTTTAGCAGT CGACGGGGTG
1151 CTGTCTGTGC ACTGCCTGCA CATCTGGTCT CTAACAATGA ATCAAGTAAT
1201 TCTCTCAGCT CATGTTGCTA CAGCAGCCAG CCGGGACAGC CAAGTGGTTC
1251 GGAGAGAAAT TGCTAAAGCC CTTAGCAAAA GCTTTACGAT GCACTCACTC
1301 ACCATTCAGA TGGAATCTCC AGTTGACCAG GACCCCGACT GCCTTTTCTG
1351 TGAAGACCCC TGTGACTAGC TCAGTCACAC CGTCAGTTTC CCAAATTTGA
1401 CAGGCCACCT TCAAACATGC TGCTATGCAA TTTCTGCATC ATAGAAAATA
1451 AGGAACCAAA GGAAGAAATT CATGTCATGG TGCAATGCAT ATTTTATCTA
1501 TTTATTTAGT TCCATTCACC ATGAAGGAAG AGGCACTGAG ATCCATCAAT
1551 CAATTGGATT ATATACTGAT CAGTAGCTGT GTTCAATTGC AGGAATGTGT
1601 ATATAGATTA TTCCTGAGTG GAGCCGAAGT AACAGCTGTT TGTAACTATC
1651 GGCAATACCA AATTCATCTC CCTTCCAATA ATGCATCTTG AGAACACATA
1701 GGTAAATTTG AACTCAGGAA AGTCTTACTA GAAATCAGTG GAAGGGACAA
1751 ATAGTCACAA AATTTACCA AAACATTAGA AACAAAAAAT AAGGAGAGCC
1801 AAGTCAGGAA TAAAAGTGAC TCTGTATGCT AACGCCACAT TAGAACTTGG
```

```
1851 TTCTCTCACC AAGCTGTAAT GTGATTTTTT TTTCTACTCT GAATTGGAAA
1901 TATGTATGAA TATACAGAGA AGTGCTTACA ACTAATTTTT ATTTACTTGT
1951 CACATTTTGG CAATAAATCC CTCTTATTTC TAAATTCTAA CTTGTTTATT
2001 TCAAAACTTT ATATAATCAC TGTTCAAAAG GAAATATTTT CACCTACCAG
2051 AGTGCTTAAA CACTGGCACC AGCCAAAGAA TGTGGTTGTA GAGACCCAGA
2101 AGTCTTCAAG AACAGCCGAC AAAAACATTC GAGTTGACCC CACCAAGTTG
2151 TTGCCACAGA TAATTTAGAT ATTTACCTGC AAGAAGGAAT AAAGCAGATG
2201 CAACCAATTC ATTCAGTCCA CGAGCATGAT GTGAGCACTG CTTTGTGCTA
2251 GACATTGGGC TTAGCACTGA AACTATAAAG AGGAATCAGA CGCAGCAAGT
2301 GCTTCTGTGT TCTGGTAGCA ACTCAACACT ATCTGTGGAG AGTAAACTGA
2351 AGATGTGCAG GCCAACATTC TGGAAATCCT ATGTCAGTGG GTTTGGTTTG
2401 GAACCTGGAC TTCTGCATTT TTAAAAGTTA CCCAGAGATG CTTCTAAAGA
2451 TGAGCCATAG TCTAGAAGAT TGTCAACCAC AGGAGTTCAT TGAGTGGGAC
2501 AGCTAGACAC ATACATTGGC AGTTACAATA GTATCATGAA TTGCAATGAT
2551 GTAGTGGGGT ATAAAAGGAA AGCGATGGAT ATTGCCGGAT GGGCATGGCC
2601 AGTGATGTTT CACGTCATTG AGGTGACAGC TCTGCTGGAC TTTGAATTAC
2651 ATATGGAGGC TCTCCAGGAA GACGAAGAAG AGAAGGACAT TCTAGGCAAA
2701 AAGAAGACTA GGCACAAGGC ACACTTATGT TTGTCTGTTA GCTTTTAGTT
2751 GAAAAAGCAA AATACATGAT GCAAAGAAAC CTCTCCACGC TGTGATTTTT
2801 AAAACTACAT ACTTTTTGCA ACTTTATGGT TATGAGTATT GTAGAGAACA
2851 GGAGATAGGT CTTAGATGAT TTTTATGTTG TTGTCAGACT CTAGCAAGGT
2901 ACTAGAAACC TAGCAGGCAT TAATAATTGT TGAGGCAATG ACTCTGAGGC
2951 TATATCTGGG CCTTGTCATT ATTTATCATT TATATTTGTA TTTTTTTCTG
3001 AAATTTGAGG GCCAAGAAAA CATTGACTTT GACTGAGGAG GTCACATCTG
3051 TGCCATCTCT GCAAATCAAT CAGCACCACT GAAATAACTA CTTAGCATTC
3101 TGCTGAGCTT TCCCTGCTCA GTAGAGACAA ATATACTCAT CCCCCACCTC
3151 AGTGAGCTTG TTTAGGCAAC CAGGATTAGA GCTGCTCAGG TTCCCAACGT
3201 CTCCTGCCAC ATCGGGTTCT CAAAATGGAA AGAATGGTTT ATGCCAAATC
3251 ACTTTTCCTG TCTGAAGGAC CACTGAATGG TTTTGTTTTT CCATATTTTG
3301 CATAGGACGC CCTAAAGACT AGGTGACTTG GCAAACACAC AAGTGTTAGT
3351 ATAATTCTTT GCTTCTGCTT CTTTTTGAAA ATCATGTTTA GATTTGATTT
3401 TAAGTCAGAA ATTCACTGAA TGTCAGGTAA TCATTATGGA GGGAGATTTG
3451 TGTGTCAACC AAAGTAATTG TCCCATGGCC CCAGGGTATT TCTGTTGTTT
3501 CCCTGAAATT CTGCTTTTTT AGTCAGCTAG ATTGAAAACT CTGAACAGTA
3551 GATGTTTATA TGGCAAAATG CAAGACAATC TATAAGGGAG ATTTTAAGGA
3601 TTTTGAGATG AAAAAACAGA TGCTACTCAG GGGCTTTATG GACCATCCAT
3651 CAATTCTGAA GTTCTGACTC TCCCATTACC CTTTCCCTGG TGTGGTCAGA
3701 ACTCCAGGTC ACTGGAAGTT AGTGGAATCA TGTAGTTGAA TTCTTTACTT
3751 CAAGACATTG TATTCTCTCC AGCTATCAAA ACATTAATGA TCTTTTATGT
3801 CTTTTTTTTG TTATTGTTAT ACTTTAAGTT CTGGGGTACA TGTGCGGAAC
3851 ATGTAGGTTT GTTACATAGG TATACATGTG CCATGGTGGT TTGCTGCACT
3901 CATCAACCTG TCATCTACAT TCTTTTATGT CTGTCTTTCA AAGCAACACT
3951 CTGTTCTTCT GAGTAGTGAA ATCAGGTCAA CTTTACCACC AGCCTCCATT
4001 TTTAATATGC TTCACCATCA TCCAGCACCT ACTTAAGATT TATCTAGGGC
4051 TCTGTGGTGA TGTTAGGACC CATAAAAGAA ATTTATGCCT TCCATATGTT
4101 TGGTTACAGA TGGGAAATGG GAATGTTGAA GGACATGAAA GAAAGGATGT
4151 TTACACATTA AGCATCAGTT CTGAAGCTAG ATTGTCTGAG TTTGAATCTT
4201 AGCTCTTCCC TTTATTAGCT CTGTGACCTC GAGCTAGTTA CTTAAATGCT
4251 CTGATCCTCT ATTTCCTGAT CAGTGAAACC TCCCTATTCA AATGTGTGAG
4301 AGTTTAATAA ATTAGGACAC TTAAAAATGT TGGAGCAGTG CATAGCATGT
4351 AGTGTTCAGT ACATGTTAAA TGTTGTTTTT TATTATGTAC AAACATGTGT
4401 GGGCACAGAA TTTTAAATCA TCTCAACTTT TGAGAAATTT TGAGTTATCA
4451 ACACCGTTCC CACAAGACAG TGGCAAAATT ATTGGTGAGA ATTAAACAGC
4501 TGTTTCTCAG AGGAAGCAAT GGAGGCTTGC TGGGATAAAG GCATTTACTG
4551 AGAGGCTGTT ACCTAGTGAG AGTGATGAAT TAATTAAAAT AGTCGAATCC
4601 CTTTCTGACT GTCTCTGAAA GCTTCCGCTT TTATCTTTGA AGAGCAGAAT
4651 TGTCACCCCA AGGACATTTA TTAATAAAAA GAACAACTGT CCAGTGCAAT
4701 GAAGGCAAAG TCATAGGTCT CCCAAGTCTT ACCCCATTCC TGTGAAATAT
4751 CAAGTTCTTG GCTTTTCTCT GTCATGTAGC CTCAACTTTC TCCGACCGGG
4801 TGCATTTCTT TCTCTGGTTT CTAAATTGCC AGTGGCAAAT TTGGATCACT
4851 TACTTAATAT CTGTTAAATT TTGTGACCCA ACAAAGTCTT TTAGCACTGT
4901 GGTGTCAAAA AGAAAAACAC CTCCCAGGCA TATACATTTT ATAGATTCCT
4951 GGAGAATGTT GCTCTCCAGC TCCATCCCCA CCCAATGAAA TATGATCCAG
5001 AGAGTCTTGC AAAGAGACAA GCCTCATTTT CCACAATTAG CTCTAAAGTG
5051 CCTCCAGGAA ATGATTTTCT CAGCTCATCT CTCTGTATTC CCTGTTTTGG
5101 ATCACAGGGC AATCTGTTTA AATGACTAAT TACAGAAATC ATTAAAGGCA
5151 CCAAGCAAAT GTCATCTCTG AATACACACA TCCCAAGCTT TACAAATCCT
5201 GCCTGGCTTG ACAGTGATGA GGCCACTTAA CAGTCCAGCG CAGGCGGATG
5251 TTAAAAAAAA TAAAAAGGTG ACCATCTGCG GTTTAGTTTT TTAACTTTCT
5301 GATTTCACAC TTAACGTCTG TCATTCTGTT ACTGGGCACC TGTTTAAATT
5351 CTATTTTAAA ATGTTAATGA GTGTTGTTTA AAATAAAATC AGGAAAGAGA
5401 GAAAAAAAAA AAAAAAAAAA AC
```

BLAST Results

No BLAST result

Medline entries

97121493:
ZnT-3, a putative transporter of zinc into synaptic vesicles.

96203098:
ZnT-2, a mammalian protein that confers resistance to zinc by facilitating vesicular sequestration.

Peptide information for frame 2

ORF from 407 bp to 1366 bp; peptide length: 320
Category: strong similarity to known protein

```
  1 MYHCHSGSKP TEKGANEYAY AKWKLCSASA ICFIFMIAEV VGGHIAGSLA
 51 VVTDAAHLLI DLTSFLLSLF SLWLSSKPPS KRLTFGWHRA EILGALLSIL
101 CIWVVTGVLV YLACERLLYP DYQIQATVMI IVSSCAVAAN IVLTVVLHQR
151 CLGHNHKEVQ ANASVRAAFV HAPGDLFQSI SVLISALIIY FKPEYKIADP
201 ICTFIFSILV LASTITILKD FSILLMEGVP KSLNYSGVKE LILAVDGVLS
251 VHCLHIWSLT MNQVILSAHV ATAASRDSQV VRREIAKALS KSFTMHSLTI
301 QMESPVDQDP DCLFCEDPCD
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_62f10, frame 2

PIR:S70632 zinc transporter ZnT-2 - rat, N = 1, Score = 884, P = 1.5e-88

TREMBL:MMU76007_1 gene: "ZnT-3"; product: "ZnT-3"; Mus musculus zinc transporter ZnT-3 (ZnT-3) mRNA, complete cds., N = 1, Score = 772, P = 1.1e-76

TREMBL:HSU76010_1 gene: "ZnT-3"; product: "ZnT-3"; Human putative zinc transporter ZnT-3 (ZnT-3) mRNA, complete cds., N = 1, Score = 742, P = 1.6e-73

TREMBL:MMUZNT02_1 gene: "ZnT-3"; product: "zinc transporter"; Mus musculus zinc transporter (ZnT-3) gene, complete cds., N = 1, Score = 715, P = 1.2e-70

TREMBL:CET18D3_3 gene: "T18D3.3"; Caenorhabditis elegans cosmid T18D3, N = 1, Score = 699, P = 5.9e-69

>PIR:S70632 zinc transporter ZnT-2 - rat
Length = 359

HSPs:

Score = 884 (132.6 bits), Expect = 1.5e-88, P = 1.5e-88
Identities = 171/326 (52%), Positives = 230/326 (70%)

```
Query:    2 YHCHSGSKPTEKGANEYAYAKWKLCSASAICFIFMIAEVVGGHIAGSLAVVTDAAHLLID 61
            ++CH+          +E   A+ KL  ASAIC +FMI E++GG++A SLA++TDAAHLL D
Sbjct:   34 HYCHAQKDSGSHPNSEKQRARRKLYVASAICLVFMIGEIIGGYLAQSLAIMTDAAHLLTD 93

Query:   62 LTSFLLSLFSLWLSSKPPSKRLTFGWHRAEILGALLSILCIWVVTGVLVYLACERLLYPD 121
              S L+SLFSLW+SS+P +K + FGW RAEILGALLS+L IWVVTGVLVYLA +RL+  D
Sbjct:   94 FASMLISLFSLWVSSRPATKTMNFGWQRAEILGALLSVLSIWVVTGVLVYLAVQRLISGD 153

Query:  122 YQIQATVMIIVSSCAVAANIVLTVVLHQRCLGHNH-------KEVQANASVRAAFVHAPG 174
            Y+I+    M+I S CAVA NI++ + LHQ    GH+H          + Q N SVRAAF+H  G
Sbjct:  154 YEIKGDTMLITSGCAVAVNIIMGLALHQSGHGHSHGHSHEDSSQQQQNPSVRAAFIHVVG 213

Query:  175 DLFQSISVLISALIIYFKPEYKIADPICTFIFSILVLASTITILKDFSILLMEGVPKSLN 234
            DL QS+ VL++A IIYFKPEYK  DPICTF+FSILVL +T+TIL+D  ++LMEG PK ++
Sbjct:  214 DLLQSVGVLVAAYIIYFKPEYKYVDPICTFLFSILVLGTTLTILRDVILVLMEGTPKGVD 273

Query:  235 YSGVKELILAVDGVLSVHCLHIWSLTMNQVILSAHVATAASRDSQVVRREIAKALSKSFT 294
            ++ VK L+L+VDGV ++H LHIW+LT+ Q +LS H+A A + D+Q V +       L   F
Sbjct:  274 FTTVKNLLLSVDGVEALHSLHIWALTVAQPVLSVHIAIAQNVDAQAVLKVARDRLQGKFN 333
```

```
Query:   295 MHSLTIQMESPVDQDPDCLFCEDPCD 320
             H++TIQ+ES  +    C  C+ P +
Sbjct:   334 FHTMTIQIESYSEDMKSCQECQGPSE 359
```

Pedant information for DKFZphfbr2_62f10, frame 2

Report for DKFZphfbr2_62f10.2

```
[LENGTH]      320
[MW]          35053.51
[pI]          6.48
[HOMOL]       PIR:S70632 zinc transporter ZnT-2 - rat 3e-84
[FUNCAT]      30.02 organization of plasma membrane       [S. cerevisiae, YMR243c] 2e-16
[FUNCAT]      13.01 homeostasis of metal ions      [S. cerevisiae, YMR243c] 2e-16
[FUNCAT]      08.19 cellular import  [S. cerevisiae, YMR243c] 2e-16 ,
[FUNCAT]      11.07 detoxificaton    [S. cerevisiae, YMR243c] 2e-16
[FUNCAT]      07.04.01 metal ion transporters (cu, fe, etc.)      [S. cerevisiae, YMR243c]
2e-16
[FUNCAT]      08.04 mitochondrial transport        [S. cerevisiae, YOR316c] 3e-13
[FUNCAT]      30.16 mitochondrial organization     [S. cerevisiae, YOR316c] 3e-13
[FUNCAT]      99 unclassified proteins     [S. cerevisiae, YDR205w] 4e-07
[PIRKW]       transmembrane protein 2e-30
[PIRKW]       mitochondrial inner membrane 6e-12
[PIRKW]       mitochondrion 6e-12
[PIRKW]       membrane protein 1e-11
[SUPFAM]      zinc transporter ZnT-2 2e-30
[SUPFAM]      membrane protein czcD 1e-11
[PROSITE]     MYRISTYL       4
[PROSITE]     CAMP_PHOSPHO_SITE      1
[PROSITE]     CK2_PHOSPHO_SITE       1
[PROSITE]     PROKAR_LIPOPROTEIN     1
[PROSITE]     TYR_PHOSPHO_SITE       1
[PROSITE]     PKC_PHOSPHO_SITE       4
[PROSITE]     ASN_GLYCOSYLATION      2
[KW]          TRANSMEMBRANE  5
[KW]          LOW_COMPLEXITY     8.12 %
```

```
SEQ    MYHCHSGSKPTEKGANEYAYAKWKLCSASAICFIFMIAEVVGGHIAGSLAVVTDAAHLLI
SEG    .........................................................xxx
PRD    cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhh
MEM    .......................MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ    DLTSFLLSLFSLWLSSKPPSKRLTFGWHRAEILGALLSILCIWVVTGVLVYLACERLLYP
SEG    xxxxxxxxxxxxxxxxxxxxxx......................................
PRD    hhhhhhhhhhhhhhhhccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhc
MEM    MMMMMMMMMMMMM..........MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.......

SEQ    DYQIQATVMIIVSSCAVAANIVLTVVLHQRCLGHNHKEVQANASVRAAFVHAPGDLFQSI
SEG    ............................................................
PRD    cccccccceeeehhhhhhhhhhhhhhhhhhcccccccccccccchhhhhhhhhhhhhchhh
MEM    MMMMMMMMMMMMMMMMMMMMMMMMMM...........MMMMMMMMMMMMMMMMMMMMM...

SEQ    SVLISALIIYFKPEYKIADPICTFIFSILVLASTITILKDFSILLMEGVPKSLNYSGVKE
SEG    ............................................................
PRD    hhhhhhhhhhcccceeecchhhhhhhhhhhhchhhhhhhheeeeeccccccchhhhhh
MEM    ..MMMMMMMMMMMMMMMMMMMM......................................

SEQ    LILAVDGVLSVHCLHIWSLTMNQVILSAHVATAASRDSQVVRREIAKALSKSFTMHSLTI
SEG    ............................................................
PRD    hhhhhhceeecccceeeeeeccchhhhheeeeeccccchhhhhhhhhhhhhhhhcccccee
MEM    ............................................................

SEQ    QMESPVDQDPDCLFCEDPCD
SEG    ....................
PRD    eeecccccccccccccccccc
MEM    ....................
```

Prosite for DKFZphfbr2_62f10.2

```
PS00001    162->166   ASN_GLYCOSYLATION     PDOC00001
PS00001    234->238   ASN_GLYCOSYLATION     PDOC00001
PS00004     81->85    CAMP_PHOSPHO_SITE     PDOC00004
PS00005     11->14    PKC_PHOSPHO_SITE      PDOC00005
PS00005     75->78    PKC_PHOSPHO_SITE      PDOC00005
```

PS00005	80->83	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00007	13->21	TYR_PHOSPHO_SITE	PDOC00007
PS00008	7->13	MYRISTYL	PDOC00008
PS00008	42->48	MYRISTYL	PDOC00008
PS00008	94->100	MYRISTYL	PDOC00008
PS00008	228->234	MYRISTYL	PDOC00008
PS00013	125->136	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphfbr2_62f10.2)

DKFZphfbr2_62n10
----

group: brain derived

DKFZphfbr2_62n10 encodes a novel 541 amino acid protein with similarity to Plasmodium vivax reticulocyte-binding protein 1.

The novel protein contains one Leucine Zipper, involved in protein-protein-interaction. No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to reticulocyte-binding protein

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="13"

Insert length: 3522 bp
Poly A stretch at pos. 3503, polyadenylation signal at pos. 3479

```
   1 GGGGCGTGTT GGCGGGATTC TGAACGCTGC CATGGCTCAG ACCGTGTAGA
  51 ATGTTACATT GTCGCTCACT CTGCCCATCA CGTGCCACAT TTGCTTGGGG
 101 AAGGTACGTC AGCCTGTCAT ATGCATCAAC AACCATGTAT TTTGTTCGAT
 151 TTGTATTGAT TTGTGGTTGA AGAATAATAG CCAGTGTCCA GCTTGCAGAG
 201 TCCCCATCAC TCCTGAAAAT CCTTGCAAAG AAATTATAGG AGGAACAAGT
 251 GAAAGTGAAC CTATGCTAAG CCATACGGTC AGGAAGCATC TTCGGAAAAC
 301 TAGACTTGAA TTACTACACA AAGAATATGA GGACGAAATA GATTGTTTAC
 351 AGAAAGAAGT AGAAGAGCTT AAGAGTAAAA ATCTCAGCTT GGAGTCACAG
 401 ATCAAAGCTA TTCTGGATCC TTTAACCTTG GTGCAGGGCA ACCAAAATGA
 451 AGACAAACAT CTAGTCACAG ATAATCCAAG TATAATTAAC CCAGAAACTG
 501 TAGCAGAGTG GAAGAAAAAA CTCAGAACAG CTAATGAAAT CTATGAAAAA
 551 GTGAAAGATG ATGTGGATAA GCTAAAGGAG GCAAATAAAA AATTGAAATT
 601 GGAAAATGGT GGTCTGGTGA GGGAGAATTT ACGACTGAAG GCTGAAGTTG
 651 ATAACAGATC ACCTCAAAAG TTTGGAAGGT TTGCAGTTGC TGCTCTTCAG
 701 TCCAAAGTAG AACAGTATGA GCGTGAAACC AATCGCCTCA AGAAAGCCCT
 751 GGAACGAAGT GATAAGTATA TAGAGGAACT AGAATCTCAA GTTGCACAGC
 801 TAAAAAATTC AAGTGAAGAG AAAGAGGCTA TGAATTCCAT TTGCCAGACA
 851 GCACTTTCTG CAGATGGCAA AGGGAGCAAA GGCAGTGAGG AGGATGTGGT
 901 GTCAAAGAAT CAAGGCGATA GTGCCAGAAA GCAGCCTGGC TCATCCACCT
 951 CCAGTTCTTC TCACCTAGCG AAGCCTTCCA GCAGCAGACT GTGTGACACC
1001 AGTTCTGCAA GGCAGGAAAG TACCAGCAAA GCAGACCTTA ACTGTTCTAA
1051 GAACAAAGAC CTATATCAAG AACAGGTAGA AGTAATGTTA GATGTGACAG
1101 ATACAAGTAT GGATACTTAT TTGGAAAGAG AATGGGGGAA TAAACCAAGT
1151 GACTGTGTAC CCTACAAAGA TGAAGAACTT TATGATTTTC CAGCTCCTTG
1201 TACTCCTTTG TCCCTTAGTT GCCTTCAGCT CAGTACTCCA GAAAATAGAG
1251 AGAGCTCTGT GGTCCAAGCA GGAGGTTCCA AAAGCACTCC AAACCATCTC
1301 AGAAAATTGG TGTTTGATGA TTTTTGTGAT TCTTCAAATG TTTCTAATAA
1351 AGATTCTTCA GAAGATGATA TAAGTAGAAG TGAAAATGAG AAGAAATCAG
1401 AATGTTTTTC TTCCACAAAG ACAGGATTTT GGGACTGTTG TTCCACAAGC
1451 TATGCCCAAA ACTTAGATTT TGAAAGTTCA GAGGGGAACA CGATAGCAAA
1501 TTCTGTTGGA GAAATATCTT CAAAATTGAG TGAGAAATCA GGCTTATGTT
1551 TATCCAAAAG GTTGAATTCT ATTCGCTCTT TTGAAATGAA CCGGACAAGA
1601 ACATCCAGTG AAGCATCGAT GGATGCTGCT TACCTTGACA AAATCTCTGA
1651 GTTGGATTCA ATGATGTCAG AGTCAGACAA CAGCAAGAGC CCTTGTAATA
1701 ACGGTTTTAA GTCACTGGAT TTGGATGGGT TATCAAAGTC ATCTCAAGGC
1751 AGTGAATTTC TTGAGGAACC TGATAAGTTG GAAGAAAAAA CTGAGCTAAA
1801 CCTTTCCAAA GGTTCTCTAA CTAATGATCA GTTAGAAAAT GGAAGTGAAT
1851 GGAAACCCAC TTCTTTTTTT TCTCCTCTCT CCATCTGACC AAGAAATGAA
1901 TGAAGATTTT TCACTCCATT CCAGTTCTTG TCCAGTAACT AATGAAATCA
1951 AACCCCCAAG CTGCTTGTTT CAGACAGAGT TTTCCCAGGG CATTTTGTTA
2001 AGCAGTTCAC ATCGACTATT GGAAGATCAA AGATTTGGGT CATCTTTGTT
2051 TAAGATGTCC TCAGAGATGC ACAGTCTTCA TAACCACCTT CAGTCTCCTT
2101 GGTCTACTTC CTTTGTGCCT GAAAAGAGGA ATAAAAATGT GAATCAATCA
2151 ACAAAAAGAA AAATCCAGAG CAGCCTTTCC AGTGCCAGCC CATCAAAAGC
2201 AACTAAAAGT TGACTCATTA GAAAGGTGTC ATTTGTGGTT TTGTCCTGAG
2251 AGAAATAGAA AAGTTGTTAA AGTTACCTTT TTTCCTCATA AAAGTTCTAT
2301 ACAAATTGGA ATTGATAATC TTTAGTCAAG TATCAAGTCA GGATGGTGGA
2351 TTAACCTGTA CCCAGAATAC TTATTGTTCA TTTTGAAAAG ACTTTGTTCT
2401 TTTCATTTTT ATTTGGGAGT CTTTGTGACC AGAGAAGTTA GGGAGGAGGT
2451 TATTTTTGTG TTTTGGGGTT GGTTGGTTGG TTGGTTTTGT TTTTGGTTTT
2501 GTTTTTTTAC TGAATTTGAT ATGTATCTCG GTTGGATATA CATTGTTTTT
2551 TTAAAAAATG TTATTTAACT GTTAGATACA GTGGCCTGTT GATAAGCCCC
2601 ACTTGTCTTC AGAACTTGGA TTTCTTAAAT AAAACTTTTA GTGTTGTCTA
```

```
2651 TACACTGCTC AATAAGACAC TTGAGTTTAA GCTTTTCCCA GGGTGGAAAT
2701 TATTTTACCT GTCCCTTTTT ATTTATGTTT AGTGATGGCC TAGTTTTTCT
2751 GCAGGGCCAT GATGGAGAAA TAGCACTCTA GCCTTAGTCC AATATTGATT
2801 TACTTTCTTT TTTTAGGTTT TATGTATATG TTTGCATTTT TTAGCATTGT
2851 GTTTTGTCCA GTTTTGTGAA AATGTTCTGC TAGTATGAAA GAAAACATTT
2901 TCTATATGAA GACATTTGTT TTATGTTAGG TAGCTTACAT TTTCTCCTCT
2951 GCGTGTGTGT GTATGTGTGT AAAATCAGAA ATTTAGCATA CTATGGAAAG
3001 AAGGCATGGA GCACTTGGGT TTAGAGGAAC CTAAAACATC ATAGCTTCAT
3051 TGTTCCAGAT GTAACAGGTT TGAAAGAGCT CATCGCCAAG TTCTTGATCC
3101 ACTTGCATTC CAGGGGAGTT CTCTTTTGAG TAGTATGTTT CTTGTTTGCA
3151 TGTTCCTGTT CTTTGTGGAA ACTATGCATG GTAGCATTTT TGCTTGCTGT
3201 GTTTTCCATA CTTAAGAAAA AGAGGTTTCA GTTGGCTGAT AGAATATCTT
3251 TTATGTAGGA CAAAACTTTT CTGTGAAGAG TGTTGAGGGG GTGAAGATAG
3301 GTAAGAGGTA AGCACAATTT TTAATTTAGG CTCTGAAAAA GTGTATTGTT
3351 CTAAACGTAT TTGGTATGCC TATATAGGTC TTTAAAAATG GGTTTGTATG
3401 CTGTTTAATG TGCACTGAAC ATTTTACATT AATATTGTAC TGTTTTACAT
3451 TAATACTGCA TGCTTTTCTA TGTGAATTGA ATAAAGAATG TCATAAGCAC
3501 TGGAAAAAAA AAAAAAAAAA AA
```

BLAST Results

Entry HS658254 from database EMBL:
human STS SHGC-11774.
Score = 1643, P = 8.0e-67, identities = 345/355

Entry HS513217 from database EMBL:
human STS SHGC-14656.
Score = 1193, P = 5.8e-46, identities = 241/244

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 263 bp to 1885 bp; peptide length: 541
Category: similarity to known protein

```
  1 MLSHTVRKHL RKTRLELLHK EYEDEIDCLQ KEVEELKSKN LSLESQIKAI
 51 LDPLTLVQGN QNEDKHLVTD NPSIINPETV AEWKKKLRTA NEIYEKVKDD
101 VDKLKEANKK LKLENGGLVR ENLRLKAEVD NRSPQKFGRF AVAALQSKVE
151 QYERETNRLK KALERSDKYI EELESQVAQL KNSSEEKEAM NSICQTALSA
201 DGKGSKGSEE DVVSKNQGDS ARKQPGSSTS SSSHLAKPSS SRLCDTSSAR
251 QESTSKADLN CSKNKDLYQE QVEVMLDVTD TSMDTYLERE WGNKPSDCVP
301 YKDEELYDFP APCTPLSLSC LQLSTPENRE SSVVQAGGSK KHSNHLRKLV
351 FDDFCDSSNV SNKDSSEDDI SRSENEKKSE CFSSTKTGFW DCCSTSYAQN
401 LDFESSEGNT IANSVGEISS KLSEKSGLCL SKRLNSIRSF EMNRTRTSSE
451 ASMDAAYLDK ISELDSMMSE SDNSKSPCNN GFKSLDLDGL SKSSQGSEFL
501 EEPDKLEEKT ELNLSKGSLT NDQLENGSEW KPTSFFSPLS I
```

BLASTP hits

Entry A42771 from database PIR:
reticulocyte-binding protein 1 - Plasmodium vivax
Score = 127, P = 3.7e-08, identities = 68/300, positives = 145/300

Entry RBP1_PLAVB from database SWISSPROT:
RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
Score = 127, P = 3.9e-08, identities = 68/300, positives = 145/300

Entry MMDSPPG_1 from database TREMBL:
gene: "DSPP"; product: "dentin sialophosphoprotein"; Mus musculus DSPP gene
Score = 160, P = 5.2e-08, identities = 87/373, positives = 146/373

Alert BLASTP hits for DKFZphfbr2_62n10, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_62n10, frame 2

Report for DKFZphfbr2_62n10.2

```
[LENGTH]      541
[MW]          60533.06
[pI]          5.10
[FUNCAT]      04.99 other transcription activities  [S. cerevisiae, YKR092c] 3e-05
[FUNCAT]      30.10 nuclear organization   [S. cerevisiae, YKR092c] 3e-05
[PROSITE]     LEUCINE_ZIPPER 1
[PROSITE]     MYRISTYL      7
[PROSITE]     CAMP_PHOSPHO_SITE     1
[PROSITE]     CK2_PHOSPHO_SITE      18
[PROSITE]     PROKAR_LIPOPROTEIN    1
[PROSITE]     TYR_PHOSPHO_SITE      1
[PROSITE]     PKC_PHOSPHO_SITE      14
[PROSITE]     ASN_GLYCOSYLATION     7
[KW]          All_Alpha
[KW]          LOW_COMPLEXITY     9.24 %
[KW]          COILED_COIL       22.55 %
```

```
SEQ    MLSHTVRKHLRKTRLELLHKEYEDEIDCLQKEVEELKSKNLSLESQIKAILDPLTLVQGN
SEG    ............................................................
PRD    ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhcccccccccc
COILS  ........CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC........

SEQ    QNEDKHLVTDNPSIINPETVAEWKKKLRTANEIYEKVKDDVDKLKEANKKLKLENGGLVR
SEG    ...................................xxxxxxxxxxxxxxxxxxxx.....
PRD    cccceeeecccccccccchhhhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhcccceee
COILS  ................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ    ENLRLKAEVDNRSPQKFGRFAVAALQSKVEQYERETNRLKKALERSDKYIEELESQVAQL
SEG    ............................................................
PRD    ehhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  ........................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ    KNSSEEKEAMNSICQTALSADGKGSKGSEEDVVSKNQGDSARKQPGSSTSSSSHLAKPSS
SEG    ..............................................xxxxxxxxxxxxxx
PRD    hcchhhhhhhhhhhhhhhcccccccccccceeeeecccccccccccccccccccccccccc
COILS  CCCCCC......................................................

SEQ    SRLCDTSSARQESTSKADLNCSKNKDLYQEQVEVMLDVTDTSMDTYLEREWGNKPSDCVP
SEG    x...........................................................
PRD    ccccccccccccccccccccccccccchhhhhhhhhccccccccccchhhhhhcccccccc
COILS  ............................................................

SEQ    YKDEELYDFPAPCTPLSLSCLQLSTPENRESSVVQAGGSKKHSNHLRKLVFDDFCDSSNV
SEG    ............................................................
PRD    cccccccccccccccccceeeecccccccccceeeeeccccccccccccccccccccccc
COILS  ............................................................

SEQ    SNKDSSEDDISRSENEKKSECFSSTKTGFWDCCSTSYAQNLDFESSEGNTIANSVGEISS
SEG    ............................................................
PRD    cccccccchhhhcccccccccccccccccccccccccccccccccccccccccccccccc
COILS  ............................................................

SEQ    KLSEKSGLCLSKRLNSIRSFEMNRTRTSSEASMDAAYLDKISELDSMMSESDNSKSPCNN
SEG    ............................................................
PRD    ccccccccchhhhcccccccccccccchhhhhhhhhhhhhhhhccccccccccccccccc
COILS  ............................................................

SEQ    GFKSLDLDGLSKSSQGSEFLEEPDKLEEKTELNLSKGSLTNDQLENGSEWKPTSFFSPLS
SEG    ..xxxxxxxxxxxxxxx...........................................
PRD    cccccccccccccccccceeecccchhhhhhhhccccccccccccccccccccccccccc
COILS  ............................................................

SEQ    I
SEG    .
PRD    c
COILS  .
```

Prosite for DKFZphfbr2_62n10.2

PS00001	40->44	ASN_GLYCOSYLATION	PDOC00001
PS00001	182->186	ASN_GLYCOSYLATION	PDOC00001
PS00001	260->264	ASN_GLYCOSYLATION	PDOC00001

PS00001	359->363	ASN_GLYCOSYLATION	PDOC00001
PS00001	443->447	ASN_GLYCOSYLATION	PDOC00001
PS00001	513->517	ASN_GLYCOSYLATION	PDOC00001
PS00001	526->530	ASN_GLYCOSYLATION	PDOC00001
PS00004	340->344	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	156->159	PKC_PHOSPHO_SITE	PDOC00005
PS00005	166->169	PKC_PHOSPHO_SITE	PDOC00005
PS00005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS00005	240->243	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	361->364	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	419->422	PKC_PHOSPHO_SITE	PDOC00005
PS00005	423->426	PKC_PHOSPHO_SITE	PDOC00005
PS00005	431->434	PKC_PHOSPHO_SITE	PDOC00005
PS00005	436->439	PKC_PHOSPHO_SITE	PDOC00005
PS00006	13->17	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	183->187	CK2_PHOSPHO_SITE	PDOC00006
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00006	281->285	CK2_PHOSPHO_SITE	PDOC00006
PS00006	285->289	CK2_PHOSPHO_SITE	PDOC00006
PS00006	324->328	CK2_PHOSPHO_SITE	PDOC00006
PS00006	361->365	CK2_PHOSPHO_SITE	PDOC00006
PS00006	365->369	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006
PS00006	373->377	CK2_PHOSPHO_SITE	PDOC00006
PS00006	414->418	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	462->466	CK2_PHOSPHO_SITE	PDOC00006
PS00006	469->473	CK2_PHOSPHO_SITE	PDOC00006
PS00007	294->302	TYR_PHOSPHO_SITE	PDOC00007
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	226->232	MYRISTYL	PDOC00008
PS00008	292->298	MYRISTYL	PDOC00008
PS00008	408->414	MYRISTYL	PDOC00008
PS00008	427->433	MYRISTYL	PDOC00008
PS00008	489->495	MYRISTYL	PDOC00008
PS00008	517->523	MYRISTYL	PDOC00008
PS00013	310->321	PROKAR_LIPOPROTEIN	PDOC00013
PS00029	104->126	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_62n10.2)

DKFZphfbr2_62o17
----

group: metabolism

DKFZphfbr2_62o17.2 encodes a novel 282 amino acid protein with weak similarity to the apolipoprotein E receptor.

The new protein contains a leucine zipper for protein-protein interaction, and three LDL-receptor class A domain (LDLRA_1) patterns. In LDL-receptors the class A domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of positively charged sequences in LDLR's ligands.

The new protein can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins

similarity to apolipoprotein E receptor

complete cDNA, complete cds, start at Bp 56 matches kozak consensus
ANCatg EST hits

Sequenced by LMU

Locus: unknown

Insert length: 1260 bp
Poly A stretch at pos. 1240, polyadenylation signal at pos. 1218

```
   1 GGGGGATAAG AGAGCGGTCT GGACAGCGCG TGGCCGGCGC CGCTGTGGGG
  51 ACAGCATGAG CGGCGGTTGG ATGGCGCAGG TTGGAGCGTG GCGAACAGGG
 101 GCTCTGGGCC TGGCGCTGCT GCTGCTGCTC GGCCTCGGAC TAGGCCTGGA
 151 GGCCGCCGCG AGCCCGCTTT CCACCCCGAC CTCTGCCCAG GCCGCAGGCC
 201 CCAGCTCAGG CTCGTGCCCA CCCACCAAGT TCCAGTGCCG CACCAGTGGC
 251 TTATGCGTGC CCCTCACCTG GCGCTGCGAC AGGGACTTGG ACTGCAGCGA
 301 TGGCAGCGAT GAGGAGGAGT GCAGGATTGA GCCATGTACC CAGAAAGGGC
 351 AATGCCCACC GCCCCCTGGC CTCCCCTGCC CCTGCACCGG CGTCAGTGAC
 401 TGCTCTGGGG GAACTGACAA GAAACTGCGC AACTGCAGCC GCCTGGCCTG
 451 CCTAGCAGGC GAGCTCCGTT GCACGCTGAG CGATGACTGC ATTCCACTCA
 501 CGTGGCGCTG CGACGGCCAC CCAGACTGTC CCGACTCCAG CGACGAGCTC
 551 GGCTGTGGAA CCAATGAGAT CCTCCCGGAA GGGGATGCCA CAACCATGGG
 601 GCCCCCTGTG ACCCTGGAGA GCGTCACCTC TCTCAGGAAT GCCACAACCA
 651 TGGGGCCCCC TGTGACCCTG GAGAGTGTCC CCTCTGTCGG GAATGCCACA
 701 TCCTCCTCTG CCGGAGACCA GTCTGGAAGC CCAACTGCCT ATGGGGTTAT
 751 TGCAGCTGCT GCGGTGCTCA GTGCAAGCCT GGTCACCGCC ACCCTCCTCC
 801 TTTTGTCCTG GCTCCGAGCC CAGGAGCGCC TCCGCCCACT GGGGTTACTG
 851 GTGGCCATGA AGGAGTCCCT GCTGCTGTCA GAACAGAAGA CCTCGCTGCC
 901 CTGAGGACAA GCACTTGCCA CCACCGTCAC TCAGCCCTGG GCGTAGCCGG
 951 ACAGGAGGAG AGCAGTGATG CGGATGGGTA CCCGGGCACA CCAGCCCTCA
1001 GAGACCTGAG CTCTTCTGGC CACGTGGAAC CTCGAACCCG AGCTCCTGCA
1051 GAAGTGGCCC TGGAGATTGA GGGTCCCTGG ACACTCCCTA TGGAGATCCG
1101 GGGAGCTAGG ATGGGGAACC TGCCACAGCC AGAACCGAGG GGCTGGCCCC
1151 AGGCAGCTCC CAGGGGGTAG GACGGCCCTG TGCTTAAGAC ACTCCTGCTG
1201 CCCCGTCTGA GGGTGGCGAT TAAAGTTGCT TCACATCCTC AAAAAAAAAA
1251 AAAAAAAAAC
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 56 bp to 901 bp; peptide length: 282
Category: similarity to known protein
Classification: unset
Prosite motifs: LDLRA_1 (67-90)
LDLRA_1 (67-90)
LDLRA_1 (145-168)

LEUCINE_ZIPPER (17-39)

```
  1 MSGGWMAQVG AWRTGALGLA LLLLLGLGLG LEAAASPLST PTSAQAAGPS
 51 SGSCPPTKFQ CRTSGLCVPL TWRCDRDLDC SDGSDEEECR IEPCTQKGQC
101 PPPPGLPCPC TGVSDCSGGT DKKLRNCSRL ACLAGELRCT LSDDCIPLTW
151 RCDGHPDCPD SSDELGCGTN EILPEGDATT MGPPVTLESV TSLRNATTMG
201 PPVTLESVPS VGNATSSSAG DQSGSPTAYG VIAAAAVLSA SLVTATLLLL
251 SWLRAQERLR PLGLLVAMKE SLLLSEQKTS LP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_62o17, frame 2

TREMBL:AF110520_6 product: "NG29"; Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial cds; and unknown gene., N = 1, Score = 733, P = 1.5e-72

PIR:JE0237 apolipoprotein E receptor 2 precursor - mouse, N = 2, Score = 290, P = 1.1e-26

TREMBL:HSZ75190_1 product: "apolipoprotein E receptor 2 906"; H.sapiens mRNA for apolipoprotein E receptor 2, N = 1, Score = 279, P = 1.8e-23

>TREMBL:AF110520_6 product: "NG29"; Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial cds; and unknown gene.
Length = 260

HSPs:

Score = 733 (110.0 bits), Expect = 1.5e-72, P = 1.5e-72
Identities = 157/276 (56%), Positives = 178/276 (64%)

```
Query:    6 MAQVGAWRTGALGLALLLLLGLGLGLEAAASPLSTPTSAQAAGPSSGSCPPTKFQCRTSG 65
            MA+ GA R  ALGL L LL GL   GLEAA +P  T     Q +G  + SCP   FQC TSG
Sbjct:    1 MARGGAGRAVALGLVLRLLFGLRTGLEAAPAPAHT--RVQVSGSRADSCPTDTFQCLTSG 58

Query:   66 LCVPLTWRCDRDLDCSDGSDEEECRIEPCTQKGQCPPPPGLPCPCTGVSDCSGGTDKKLR 125
             CVPL+WRCD D DCSDGSDEE+CRIE C Q GQC P    LPC C   +S CS   +DK L
Sbjct:   59 YCVPLSWRCDGDQDCSDGSDEEDCRIESCAQNGQCQPQSALPCSCDNISGCSDVSDKNL- 117

Query:  126 NCSRLACLAGELRCTLSDDCIPLTWRCDGHPDCPDSSDELGCGTNEILPEGDATTMGPPV 185
            NCSR  C    EL C L D CIP TWRCDGHPDC DSSDEL C T+
Sbjct:  118 NCSRPPCQESELHCILDDVCIPHTWRCDGHPDCLDSSDELSCDTD--------------T 163

Query:  186 TLESVTSLRNATTMGPPVTLESVPSVGNATSSSAGDQSGSPTAYGVIAAAAVLSASLVTA 245
              ++ +     NATT       T+E+   S   N T +SAGD S +P+AYGVIAAA VLSA LV+A
Sbjct:  164 EIDKIFQEENATTTRISTTMENETSFRNVTFTSAGDSSRNPSAYGVIAAAGVLSAILVSA 223

Query:  246 TLLLLSWLRAQERLRPLGLLVAMKESLLLSEQKTSL 281
            TLL+L   LR Q   L P GLLVA+KESLLLSE+KTSL
Sbjct:  224 TLLILLRLRGQGYLPPPGLLVAVKESLLLSERKTSL 259
```

Pedant information for DKFZphfbr2_62o17, frame 2

Report for DKFZphfbr2_62o17.2

[LENGTH] 282
[MW] 28991.19
[pI] 4.61
[HOMOL] TREMBL:AF110520_6 product: "NG29"; Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial cds; and unknown gene. 5e-55
[BLOCKS] BL01209 LDL-receptor class A (LDLRA) domain proteins
[SCOP] d1ajj__ 7.11.1.1.1 Ligand-binding domain of low-density lipoprotei 2e-10

```
[PIRKW]        duplication 1e-19
[PIRKW]        tandem repeat 1e-15
[PIRKW]        heterodimer 6e-18
[PIRKW]        endocytosis 4e-18
[PIRKW]        heparan sulfate 2e-12
[PIRKW]        VLDL 1e-19
[PIRKW]        transmembrane protein 1e-19
[PIRKW]        coated pits 4e-18
[PIRKW]        fatty acid metabolism 1e-19
[PIRKW]        G protein-coupled receptor 1e-10
[PIRKW]        receptor 1e-19
[PIRKW]        glycoprotein 1e-19
[PIRKW]        lipid transport 4e-18
[PIRKW]        LDL 5e-14
[PIRKW]        calcium binding 6e-18
[PIRKW]        extracellular protein 6e-13
[PIRKW]        alternative splicing 1e-19
[PIRKW]        extracellular matrix 3e-10
[PIRKW]        chondroitin sulfate proteoglycan 2e-12
[PIRKW]        cholesterol 4e-18
[SUPFAM]       leucine-rich alpha-2-glycoprotein repeat homology 1e-10
[SUPFAM]       LDL receptor YWTD-containing repeat homology 1e-19
[SUPFAM]       trypsin homology 6e-13
[SUPFAM]       alpha-2-macroglobulin receptor 6e-18
[SUPFAM]       LDL receptor 1e-19
[SUPFAM]       LDL receptor ligand-binding repeat homology 1e-19
[SUPFAM]       EGF homology 1e-19
[PROSITE]      LDLRA_1 3
[PROSITE]      LEUCINE_ZIPPER 1
[PFAM]         Low-density lipoprotein receptor domain class A
[PFAM]         TNFR/NGFR cysteine-rich region
[KW]           SIGNAL_PEPTIDE 31
[KW]           TRANSMEMBRANE  1
[KW]           LOW_COMPLEXITY    22.34 %
```

```
SEQ   MSGGWMAQVGAWRTGALGLALLLLLGLGLGLEAAASPLSTPTSAQAAGPSSGSCPPTKFQ
SEG   ..............xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.............
PRD   cccccccccccchhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccceee
MEM   ............................................................

SEQ   CRTSGLCVPLTWRCDRDLDCSDGSDEEECRIEPCTQKGQCPPPPGLPCPCTGVSDCSGGT
SEG   .......................................xxxxxxxxxxx..........
PRD   ecccccceeeeecccccccccccccccccccccccccccccccccccccccccccccccc
MEM   ............................................................

SEQ   DKKLRNCSRLACLAGELRCTLSDDCIPLTWRCDGHPDCPDSSDELGCGTNEILPEGDATT
SEG   ............................................................
PRD   cccccccccccccccceeeccccccccccccccccccccccccccccccccccccccccc
MEM   ............................................................

SEQ   MGPPVTLESVTSLRNATTMGPPVTLESVPSVGNATSSSAGDQSGSPTAYGVIAAAAVLSA
SEG   ....................................................xxxxxxxx
PRD   ccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhh
MEM   .....................................................MMMMMMM

SEQ   SLVTATLLLLSWLRAQERLRPLGLLVAMKESLLLSEQKTSLP
SEG   xxxxxxxxxxx...............................
PRD   hhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhcccccc
MEM   MMMMMMMMMM................................
```

Prosite for DKFZphfbr2_62o17.2

PS01209	67->90	LDLRA_1	PDOC00929
PS01209	67->90	LDLRA_1	PDOC00929
PS01209	145->168	LDLRA_1	PDOC00929
PS00029	17->39	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphfbr2_62o17.2

```
HMM_NAME     TNFR/NGFR cysteine-rich region

HMM             *CpeGtYtD.WNHvpqClpC.trCePEMGQYMvqPCTwTQNT.VC*
                 CP+  ++  +   + C+P   RC+   ++     +C + ++  +C
Query        54  CPPTKFQCRTS--GLCVPLTWRCDR--DL----DCSDGSDEEEC     89
```

```
HMM_NAME        Low-density lipoprotein receptor domain class A

HMM                 *tTCeGPDEFQCgSGeMRCIPMsWvCDGDpDCeDWSDEWPeNChp*
                      C  P +FQC+++    C+P+ W+CD D DC D+SDE  E+C+
Query          52  GSCP-PTKFQCRTSG-LCVPLTWRCDRDLDCSDGSDE--EECRI     91

54.99  (bits) f:  130 t:  169 Target: dkfzphfbr2_62o17.2 similarity to apolipoprotein E
receptor
  Alignment to HMM consensus:
Query               *tTCeGPDEFQCgSGeMRCIPMsWvCDGDpDCeDWSDEWPeNChp*
                      C  + E +C  +    CIP+ W+CDG PDC D SDE  ++C+
  dkfzphfbr2   130  LACL-AGELRCTLSD-DCIPLTWRCDGHPDCPDSSDE--LGCGT     169
```

DKFZphfbr2_64a15
----

group: nucleic acid management

DKFZphfbr2_64a15 encodes a novel 255 amino acid protein with strong similarity to inorganic pyrophosphatases

Inorganic pyrophosphatase (EC 3.6.1.1) (PPase) is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed as the product of the many biosynthetic reactions that utilize ATP. All known PPases require the presence of divalent metal cations, with magnesium conferring the highest activity.

The new protein can find application as a new enzyme for biotechnologic processes.

strong similarity to inorganic pyrophosphatases

unspliced Intron 212-256 see EST HS1190948

Sequenced by Qiagen

Locus: unknown

Insert length: 1188 bp
Poly A stretch at pos. 1170, polyadenylation signal at pos. 1151

```
   1 GGGGGTTGGG GACCAGTGCA GGGACCGGGT CGCGCCGTGC TATGGCCCTG
  51 TACCACACTG AGGAGCGCGG CCAGCCCTGC TCGCAGAATT ACCGCCTCTT
 101 CTTTAAGAAT GTAACTGGTC ACTACATTTC CCCCTTTCAT GATATTCCTC
 151 TGAAGGTGAA CTCTAAAGAG GACACTGAGG CTCAAGGCAT TTTTATAGAC
 201 TTGTCTAAGA TCTGGAAAAT GGCATTCCTA TGAAGAAAGC ACGAAATGAT
 251 GAATATGAGA ATCTGTTTAA TATGATTGTA GAAATACCTC GGTGGACAAA
 301 GGCTAAAATG GAGATTGCCA CCAAGGAGCC AATGAATCCC ATTAAACAAT
 351 ATGTAAAGGA TGGAAAGCTA CGCTATGTGG CGAATATCTT CCCTTACAAG
 401 GGTTATATAT GGAATTATGG TACCCTCCCT CAGACTTGGG AAGATCCCCA
 451 TGAAAAAGAT AAGAGCACGA ACTGCTTTGG AGATAATGAT CCTATTGATG
 501 TTTGCGAAAT AGGCTCAAAG ATTCTTTCTT GTGGAGAAGT TATTCATGTG
 551 AAGATCCTTG GAATTTTGGC TCTTATTGAT GAAGGTGAAA CAGATTGGAA
 601 ATTAATTGCT ATCAATGCGA ATGATCCTGA AGCCTCAAAG TTTCATGATA
 651 TTGATGATGT TAAGAAGTTC AAACCGGGTT ACCTGGAAGC TACTCTTAAT
 701 TGGTTTAGAT TATGTAAGGT ACCAGATGGA AAACCAGAAA ACCAGTTTGC
 751 TTTTAATGGA GAATTCAAAA ACAAGGCTTT TGCTCTTGAA GTTATTAAAT
 801 CCACTCATCA ATGTTGGAAA GCATTGCTTA TGAAGAACTG TAATGGAGGA
 851 GCTACAAATT GCACAAACGT GCAGATATCT GATAGCCCTT TCCGTTGCAC
 901 TCAAGAGGAA GCAAGATCAT TAGTTGAATC GGTATCATCT TCACCAAATA
 951 AAGAAAGTAA TGAAGAAGAG CAAGTGTGGC ACTTCCTTGG CAAGTGATTG
1001 AAACATCTGA AATTCTGCTG TCAAGATTCC CATCTCTAAG GACTCCAAGA
1051 CTCTTTTTCC CCAAGTGCTA GAGACAAGGG GGTCTATGAG CATTTACTGA
1101 CTTCCTGTTA AAACTTCATT TTTTCAAACT TTTTGAGCTA TGCAATATAT
1151 AAATAAACAG TAAGAATTTT AAAAAAAAAA AAAAAAAA
```

BLAST Results

Entry HSPPASEMR from database EMBL:
H.sapiens partial mRNA for pyrophosphatase.
Score = 1706, P = 1.6e-70, identities = 342/343

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 230 bp to 994 bp; peptide length: 255
Category: strong similarity to known protein
Classification: unset
Prosite motifs: PPASE (85-92)

```
  1 MKKARNDEYE NLFNMIVEIP RWTKAKMEIA TKEPMNPIKQ YVKDGKLRYV
 51 ANIFPYKGYI WNYGTLPQTW EDPHEKDKST NCFGDNDPID VCEIGSKILS
101 CGEVIHVKIL GILALIDEGE TDWKLIAINA NDPEASKFHD IDDVKKFKPG
151 YLEATLNWFR LCKVPDGKPE NQFAFNGEFK NKAFALEVIK STHQCWKALL
201 MKNCNGGATN CTNVQISDSP FRCTQEEARS LVESVSSSPN KESNEEEQVW
251 HFLGK
```

BLASTP hits

Entry IPYR_KLULA from database SWISSPROT:
INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE PHOSPHO-HYDROLASE) (PPASE).
Score = 689, P = 6.0e-68, identities = 128/248, positives = 170/248

Entry A45153 from database PIR:
inorganic pyrophosphatase (EC 3.6.1.1) - bovine
Score = 862, P = 2.8e-86, identities = 146/226, positives = 190/226

Entry AF085600_1 from database TREMBLNEW:
gene: "Nurf-38"; product: "inorganic pyrophosphatase NURF-38";
Drosophila melanogaster inorganic pyrophosphatase NURF-38 (Nurf-38)
gene, complete cds.
Score = 731, P = 2.1e-72, identities = 134/248, positives = 177/248

Entry PWBY from database PIR:
inorganic pyrophosphatase (EC 3.6.1.1) - yeast (Saccharomyces
cerevisiae)
Score = 688, P = 7.7e-68, identities = 133/251, positives = 174/251

Alert BLASTP hits for DKFZphfbr2_64a15, frame 2

SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1)
(PYROPHOSPHATE PHOSPHO- HYDROLASE) (PPASE)., N = 1, Score = 731, P =
2.4e-72

```
>SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE
     PHOSPHO- HYDROLASE) (PPASE).
            Length = 290

  HSPs:

 Score = 731 (109.7 bits), Expect = 2.4e-72, P = 2.4e-72
 Identities = 134/248 (54%), Positives = 177/248 (71%)

Query:     7 DEYENLFNMIVEIPRWTKAKMEIATKEPMNPIKQYVKDGKLRYVANIFPYKGYIWNYGTL 66
             +E + ++NM+VE+PRWT AKMEI+ K PMNPIKQ +K GKLR+VAN FP+KGYIWNYG L
Sbjct:    40 NEEKTIYNMVVEVPRWTNAKMEISLKTPMNPIKQDIKKGKLRFVANCFPHKGYIWNYGAL 99

Query:    67 PQTWEDPHEKDKSTNCFGDNDPIDVCEIGSKILSCGEVIHVKILGILALIDEGETDWKLI 126
             PQTWE+P    + ST C GDNDPIDV EIG ++   G+V+ VK+LG  ALIDEGETDWK+I
Sbjct:   100 PQTWENPDHIEPSTGCKGDNDPIDVIEIGYRVAKRGDVLKVKVLGQFALIDEGETDWKII 159

Query:   127 AINANDPEASKFHDIDDVKKFKPGYLEATLNWFRLCKVPDGKPENQFAFNGEFKNKAFAL 186
             AI+ NDP ASK +DI DV ++ PG L AT+ WF++ K+PDGKPENQFAFNG+ KN  FA
Sbjct:   160 AIDVNDPLASKVNDIADVDQYFPGLLRATVEWFKIYKIPDGKPENQFAFNGDAKNADFAN 219

Query:   187 EVIKSTHQCWKALLMKNCNGGATNCTNVQISDSPFRCTQEEARS-LVESVSSSPNKESNE 245
             +I  TH+ W+ L+ ++    G+ + TN+    +S      +EEA    L E+       +E ++
Sbjct:   220 TIIAETHKFWQNLVHQSPASGSISTTNITNRNSEHVIPKEEAEKILAEAPDGGQVEEVSD 279

Query:   246 EEQVWHFL 253
                  WHF+
Sbjct:   280 TVDTWHFI 287
```

Peptide information for frame 3

ORF from 42 bp to 230 bp; peptide length: 63
Category: strong similarity to known protein
Classification: unset

```
 1 MALYHTEERG QPCSQNYRLF FKNVTGHYIS PFHDIPLKVN SKEDTEAQGI
51 FIDLSKIWKM AFL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64a15, frame 3

SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE PHOSPHO- HYDROLASE) (PPASE)., N = 1, Score = 118, P = 8.8e-07

PIR:A45153 inorganic pyrophosphatase (EC 3.6.1.1) - bovine, N = 1, Score = 113, P = 3.1e-06

TREMBLNEW:AF108211_1 product: "cytosolic inorganic pyrophosphatase"; Homo sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds., N = 1, Score = 106, P = 1.8e-05

```
>SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE
     PHOSPHO- HYDROLASE) (PPASE).
          Length = 290

  HSPs:

 Score = 118 (17.7 bits), Expect = 8.8e-07, P = 8.8e-07
 Identities = 23/43 (53%), Positives = 29/43 (67%)

Query:      1 MALYHTEERGQPCSQNYRLFFKNVTGHYISPFHDIPLKVNSKE 43
              MALY T E+G   S +Y L+FKN  G+ ISP HDIPL  N ++
Sbjct:      1 MALYETVEKGAKNSPSYSLYFKNKCGNVISPMHDIPLYANEEK 43
```

Pedant information for DKFZphfbr2_64a15, frame 2

Report for DKFZphfbr2_64a15.2

```
[LENGTH]        255
[MW]            29177.34
[pI]            5.67
[HOMOL]         TREMBLNEW:AF108211_1 product: "cytosolic inorganic pyrophosphatase"; Homo
sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds. 2e-93
[FUNCAT]        01.04.01 phosphate utilization      [S. cerevisiae, YBR011c] 9e-73
[FUNCAT]        30.03 organization of cytoplasm      [S. cerevisiae, YBR011c] 9e-73
[FUNCAT]        02.99 other energy generation activities     [S. cerevisiae, YMR267w] 1e-58
[FUNCAT]        30.16 mitochondrial organization     [S. cerevisiae, YMR267w] 1e-58
[FUNCAT]        1 genome replication, transcription, recombination and repair      [M.
genitalium, MG351] 1e-06
[FUNCAT]        g carbohydrate metabolism and transport     [H. influenzae, HI0124] 2e-06
[BLOCKS]        BL00387D
[BLOCKS]        BL00387C
[BLOCKS]        BL00387B
[BLOCKS]        BL00387A
[SCOP]          d1wgja_ 2.29.5.1.1 Inorganic pyrophosphatase  [baker's yeas 1e-113
[EC]            3.6.1.1 Inorganic pyrophosphatase 7e-92
[PIRKW]         mitochondrion 3e-57
[PIRKW]         hydrolase 7e-92
[PIRKW]         homodimer 2e-71
[SUPFAM]        inorganic pyrophosphatase 7e-92
[PROSITE]       PPASE  1
[KW]            Alpha_Beta
[KW]            3D
[KW]            LOW_COMPLEXITY     6.27 %

SEQ    MKKARNDEYENLFNMIVEIPRWTKAKMEIATKEPMNPIKQYVKDGKLRYVANIFPYKGYI
SEG    ............................................................
1hukB  ......EGGGCEEEEEEEETTTbCBCEEETTTTTTTCEEECEETTEECBCCBBTTBTTbT

SEQ    WNYGTLPQTWEDPHEKDKSTNCFGDNDPIDVCEIGSKILSCGEVIHVKILGILALIDEGE
SEG    ............................................................
1hukB  CEEEETTTTCBTTTTEETTTTEECCCBCCEEEECCCCCCTTTEEEEEEEEEEEETTTTB

SEQ    TDWKLIAINANDPEASKFHDIDDVKKFKPGYLEATLNWFRLCKVPDGKPENQFAFNGEFK
SEG    ............................................................
1hukB  CEEEEEEEETTTTTGGGCCCHHHHHHHTTTHHHHHHHHHHHCGGGCCCCCCBCGGGCCB

SEQ    NKAFALEVIKSTHQCWKALLMKNCNGGATNCTNVQISDSPFFRCTQEEARSLVESVSSSPN
SEG    ....................................................xxxxxxxxx
1hukB  CHHHHHHHHHHHHHHHHHHHHCTTTTTTTCCCBTTTTTTT....................
```

```
SEQ     KESNEEEQVWHFLGK
SEG     xxxxxxx........
1hukB   ...............
```

Prosite for DKFZphfbr2_64a15.2

```
PS00387        85->92    PPASE                   PDOC00325
```

(No Pfam data available for DKFZphfbr2_64a15.2)

Pedant information for DKFZphfbr2_64a15, frame 3

Report for DKFZphfbr2_64a15.3

```
[LENGTH]        63
[MW]            7405.54
[pI]            6.81
[HOMOL]         SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE
PHOSPHO- HYDROLASE) (PPASE). 1e-06
[EC]            3.6.1.1 Inorganic pyrophosphatase 5e-06
[PIRKW]         hydrolase 5e-06
[SUPFAM]        inorganic pyrophosphatase 5e-06
[KW]            All_Beta
```

```
SEQ     MALYHTEERGQPCSQNYRLFFKNVTGHYISPFHDIPLKVNSKEDTEAQGIFIDLSKIWKM
PRD     ccccccccccccccceeeeeecccccccccccccccccccccccccccceeeechhhhhhh

SEQ     AFL
PRD     ccc
```

(No Prosite data available for DKFZphfbr2_64a15.3)

(No Pfam data available for DKFZphfbr2_64a15.3)

DKFZphfbr2_64c16
----

group: brain derived

DKFZphfbr2_64a16.2 encodes a novel 101 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: /map="745_A_2; 756_F_2; 842_C_2"

Insert length: 1866 bp
Poly A stretch at pos. 1848, polyadenylation signal at pos. 1829

```
   1 GGGCGCGGCG CCGGAGGAGG AAGTGGTGAG GTTGTTGCTC CTTCAGCGCC
  51 TATCGCTGGC TCTTGGGGCG CAGAGAGGGG CCGCAGTCTC CGCGGCTGCG
 101 TCGAGCTCCC TTGCAGTCCC CTCCATGTTC CCCGGCGCCA CTACTCCCCT
 151 TCCTAAGGCC GCCGCTTACC CCGGGGTCTA TGGAAGTAAT GGAAGGACCC
 201 CTCAACCTGG CTCATCAACA GAGCAGACGA GCAGACCGTT TATTAGCTGC
 251 AGGCAAATAC GAAGAGGCTA TTTCTTGTCA CAAAAAGGCT GCAGCATATC
 301 TTTCTGAAGC CATGAAGCTG ACACAGTCAG AGCAGGCTCA TCTTTCACTG
 351 GAATTGCAAA GGGATAGCCA TATGAAACAG CTCCTCCTCA TCCAAGAGAG
 401 ATGGAAAAGG GCCCAGCGTG AAGAAAGATT GAAAGCCCAG CAGAACACAG
 451 ACAAGGATGC AGCTGCCCAT CTTCAGACAT CTCACAAACC CTCTGCAGAG
 501 GATGCAGAGG GCCAGAGTCC CCTTTCTCAG AAGTACAGCC CTTCCACAGA
 551 GAAATGCCTG CCTGAGATTC AGGGGATCTT TGACAGGGAT CCAGACACAC
 601 TACTTTATTT ACTTCAGCAA AAGAGTGAGC CAGCAGAGCC ATGTATTGGA
 651 AGCAAAGCCC CAAAAGATGA TAAAACAATT ATAGAGGAGC AGGCAACCAA
 701 AATTGCAGAT TTGAAGAGGC ATGTGGAATT CCTTGTGGCT GAGAATGAAA
 751 GATTAAGGAA AGAAAATAAA CAACTAAAGG CTGAAAAGGC CAGACTTCTA
 801 AAAGGTCCAA TAGAAAAGGA GCTGGATGTA GATGCTGATT TTGTAGAAAC
 851 GTCAGAGTTA TGGAGCTTGC CACCACATGC AGAAACTGCT ACAGCCTCCT
 901 CAACCTGGCA GAAGTTCGCA GCAAATACTG GGAAAGCCAA GGACATTCCA
 951 ATCCCCAATC TTCCTCCCTT GGATTTTCCA TCTCCAGAAC TTCCTCTTAT
1001 GGAGCTCTCT GAGGATATTC TGAAAGGACT TATGAATAAT TAAAATGGAA
1051 GGCCACAGAA AAGGGGAAAA GAGGAAATAA TACAGTAATC GTTAATCCAG
1101 CAAAAAGAAA TGAAAAGGGA AAACCACATA GAAGGGTAAT CCCGGAAATG
1151 CTTCATCTGG TGGACTGTGG GAGCAGAGGC ATTGCCAGGA CTTGGGAAAC
1201 AGTCACTGTG AAATGCGCTG CGTATCTCAT TCACTCACTT CAGCTAATGA
1251 CTCCGACTTG GCAGACGCTA AACTCATGGA GGTTCGGTTT CTCCTGATAC
1301 AAACCAAATG GCTACCTGGA AGAATTTCTT TCAAGCAACA GTTATTTTTC
1351 TTATCTTCAG GGTTAAAATG TATAAAAGTT ATGTGTAATT AATCTATAAT
1401 GCCATAAATG ATAATGCAAA ACCTAAATAA TATGGTGGCC GGAGGGGCTG
1451 CCTTATATTT GAAACATGCT TTCTATCATG CATTGACTGT ATGCATTTTG
1501 TTAATGCACA TTCTGTTTGT TTAAGGTGTG TGAGATACAC ACCTTTCTAG
1551 ATGAAACTAT ATGTGCCACA CTTTGCACTA CTCATAATGA TAACCTCAAG
1601 ACTATCAGAA GAAATATTTA AATTTCCATT TTATGAAGAA AGGAACCAAA
1651 TTATTATGCT TTTTAAAACA AATTACCAGT TTACATAATT AATCAGGGTG
1701 CATTTTAAGT TCTAACTTCG TTTATTGTAT AATGCATCAT TTGAAAATAC
1751 CAAGGAGGAA ATACCCTTTG TTTTTAATGA TGCAAGAGTG GACGTAATGC
1801 TAGTTGGCAG TATTTTATTG TAAGAAATCA ATAAAGTAAT TGTGTTTTAA
1851 AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HS286143 from database EMBL:
human STS WI-6844.
Score = 1460, P = 3.4e-61, identities = 292/292

Medline entries

No Medline entry

Peptide information for frame 2

ORF from the beginning to 304 bp; peptide length: 102
Category: questionable ORF
Classification: unset

```
  1 GAAPEEEVVR LLLLQRLSLA LGAQRGAAVS AAASSSLAVP SMFPGATTPL
 51 PKAAAYPGVY GSNGRTPQPG SSTEQTSRPF ISCRQIRRGY FLSQKGCSIS
101 F
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64c16, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 180 bp to 1040 bp; peptide length: 287
Category: putative protein
Classification: unset
Prosite motifs: LEUCINE_ZIPPER (178-200)
LEUCINE_ZIPPER (185-207)

```
  1 MEVMEGPLNL AHQQSRRADR LLAAGKYEEA ISCHKKAAAY LSEAMKLTQS
 51 EQAHLSLELQ RDSHMKQLLL IQERWKRAQR EERLKAQQNT DKDAAAHLQT
101 SHKPSAEDAE GQSPLSQKYS PSTEKCLPEI QGIFDRDPDT LLYLLQQKSE
151 PAEPCIGSKA PKDDKTIIEE QATKIADLKR HVEFLVAENE RLRKENKQLK
201 AEKARLLKGP IEKELDVDAD FVETSELWSL PPHAETATAS STWQKFAANT
251 GKAKDIPIPN LPPLDFPSPE LPLMELSEDI LKGLMNN
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64c16, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_64c16, frame 2

Report for DKFZphfbr2_64c16.2

```
[LENGTH]    101
[MW]        10469.94
[pI]        10.18
[KW]        All_Alpha
[KW]        LOW_COMPLEXITY    29.70 %

SEQ   GAAPEEEVVRLLLLQRLSLALGAQRGAAVSAAASSSLAVPSMFPGATTPLPKAAAYPGVY
SEG   .........xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....................
PRD   ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccc

SEQ   GSNGRTPQPGSSTEQTSRPFISCRQIRRGYFLSQKGCSISF
SEG   .........................................
PRD   ccccccccccccccccccccchhhhhccccccccccccccc
```

(No Prosite data available for DKFZphfbr2_64c16.2)

(No Pfam data available for DKFZphfbr2_64c16.2)

Pedant information for DKFZphfbr2_64c16, frame 3

Report for DKFZphfbr2_64c16.3

```
[LENGTH]        287
[MW]            32343.79
[pI]            5.61
[PROSITE]       LEUCINE_ZIPPER 2
[KW]            All_Alpha
[KW]            COILED_COIL       14.98 %

SEQ     MEVMEGPLNLAHQQSRRADRLLAAGKYEEAISCHKKAAAYLSEAMKLTQSEQAHLSLELQ
PRD     ccccchhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ............................................................

SEQ     RDSHMKQLLLIQERWKRAQREERLKAQQNTDKDAAAHLQTSHKPSAEDAEGQSPLSQKYS
PRD     hhcchhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhccccccccccccccccccccc
COILS   ............................................................

SEQ     PSTEKCLPEIQGIFDRDPDTLLYLLQQKSEPAEPCIGSKAPKDDKTIIEEQATKIADLKR
PRD     ccccccchhhhhccccccchhhhhhhhhcccccccccccccccchhhhhhhhhhhhhhhh
COILS   ..............................................CCCCCCCCCCCCCC

SEQ     HVEFLVAENERLRKENKQLKAEKARLLKGPIEKELDVDADFVETSELWSLPPHAETATAS
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccc
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCCCC...............................

SEQ     STWQKFAANTGKAKDIPIPNLPPLDFPSPELPLMELSEDILKGLMNN
PRD     hhhhhhhhcccccccccccccccccccccchhhhhhhhhhhhhhccc
COILS   ...............................................
```

Prosite for DKFZphfbr2_64c16.3

```
PS00029     178->200    LEUCINE_ZIPPER          PDOC00029
PS00029     185->207    LEUCINE_ZIPPER          PDOC00029
```

(No Pfam data available for DKFZphfbr2_64c16.3)

DKFZphfbr2_64c4

group: brain derived

DKFZphfbr2_64c4 encodes a novel 467 amino acid protein with similarity to A. thaliana T08I13.5

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A. thaliana T08I13.5

complete cDNA, complete cds, EST hits
on genomic level encoded by AC005043 11 exons

Sequenced by Qiagen

Locus: unknown

Insert length: 1559 bp
Poly A stretch at pos. 1540, no polyadenylation signal found

```
   1 TGGGACCGCC GGAAGTTTCT GCCGCGGCTT TGCGGGGACG GGGGAGTGGT
  51 AGTGGGGGCT GCAGCTGCCG GACCCAGGCG CGATGGCTAC GGGCGCGGAT
 101 GTACGGGACA TTCTAGAACT CGGGGGTCCA GAAGGGGATG CAGCCTCTGG
 151 GACCATCAGC AAGAAGGACA TTATCAACCC GGACAAGAAA AAATCCAAGA
 201 AGTCCTCTGA GACACTGACT TTCAAGAGGC CCGAGGGCAT GCACCGGGAA
 251 GTCTATGCCT TGCTCTACTC TGACAAGAAG GATGCACCCC CACTGCTACC
 301 CAGTGACACT GGCCAGGGAT ACCGTACAGT GAAGGCCAAG TTGGGCTCCA
 351 AGAAGGTGCG GCCTTGGAAG TGGATGCCAT TCACCAACCC GGCCCGCAAG
 401 GACGGAGCAA TGTTCTTCCA CTGGCGACGT GCAGCGGAGG AGGGCAAGGA
 451 CTACCCCTTT GCCAGGTTCA ATAAGACTGT GCAGGAGCCT GTGTACTCGG
 501 AGCAGGAGTA CCAGCTTTAT CTCCACGATA ATGCTTGGAC TAAGGCAGAA
 551 ACTGACCACC TCTTTGACCT CAGCCGCCGC TTTGACCTGC GTTTTGTTGT
 601 TATCCATGAC CGGTATGACC ACCAGCAGTT CAAGAAGCGT TCTGTGGAAG
 651 ACCTGAAGGA GCGGTACTAC CACATCTGTG CTAAGCTTGC CAACGTGCGG
 701 GCTGTGCCAG GCACAGACCT TAAGATACCA GTATTTGATG CTGGGCACGA
 751 ACGACGGCGG AAGGAACAGC TTGAGCGTCT CTACAACCGG ACCCCAGAGC
 801 AGGTGGCAGA GGAGGAGTAC CTGCTACAGG AGCTGCGCAA GATTGAGGCC
 851 CGGAAGAAGG AGCGGGAGAA ACGCAGCCAG GACCTGCAGA AGCTGATCAC
 901 AGCGGCAGAC ACCACTGCAG AGCAGCGGCG CACGGAACGC AAGGCCCCCA
 951 AAAAGAAGCT ACCCCAGAAA AAGGAGGCTG AGAAGCCGGC TGTTCCTGAG
1001 ACTGCAGGCA TCAAGTTTCC AGACTTCAAG TCTGCAGGTG TCACGCTGCG
1051 GAGCCAACGG ATGAAGCTGC CAAGCTCTGT GGGACAGAAG AAGATCAAGG
1101 CCCTGGAACA GATGCTGCTG GAGCTTGGTG TGGAGCTGAG CCCGACACCT
1151 ACGGAGGAGC TGGTGCACAT GTTCAATGAG CTGCGAAGCG ACCTGGTGCT
1201 GCTCTACGAG CTCAAGCAGG CCTGTGCCAA CTGCGAGTAT GAGCTGCAGA
1251 TGCTGCGGCA CCGTCATGAG GCACTGGCCC GGGCTGGTGT GCTAGGGGGC
1301 CCTGCCACAC CAGCATCAGG CCCAGGCCCG GCCTCTGCTG AGCCGGCAGT
1351 GTCTGAACCC GGACTTGGTC CTGACCCCAA GGACACCATC ATTGATGTGG
1401 TGGGCGCACC CCTCACGCCC AATTCGAGAA AGCGACGGGA GTCGGCCTCC
1451 AGCTCATCTT CCGTGAAGAA AGCCAAGAAG CCGTGAGAGG CCCCACGGGG
1501 TGTGGGCGAC GCTGTTATGT AAATAGAGCT GCTGAGTTGG AAAAAAAAAA
1551 AAAAAAAAA
```

BLAST Results

Entry AC005043 from database EMBL:
Homo sapiens clone NH0576N21; HTGS phase 1, 5 unordered pieces.
Score = 1506, P = 4.6e-244, identities = 316/330

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 83 bp to 1483 bp; peptide length: 467

Category: similarity to unknown protein

```
  1 MATGADVRDI LELGGPEGDA ASGTISKKDI INPDKKKSKK SSETLTFKRP
 51 EGMHREVYAL LYSDKKDAPP LLPSDTGQGY RTVKAKLGSK KVRPWKWMPF
101 TNPARKDGAM FFHWRRAAEE GKDYPFARFN KTVQEPVYSE QEYQLYLHDN
151 AWTKAETDHL FDLSRRFDLR FVVIHDRYDH QQFKKRSVED LKERYYHICA
201 KLANVRAVPG TDLKIPVFDA GHERRRKEQL ERLYNRTPEQ VAEEEYLLQE
251 LRKIEARKKE REKRSQDLQK LITAADTTAE QRRTERKAPK KKLPQKKEAE
301 KPAVPETAGI KFPDFKSAGV TLRSQRMKLP SSVGQKKIKA LEQMLLELGV
351 ELSPTPTEEL VHMFNELRSD LVLLYELKQA CANCEYELQM LRHRHEALAR
401 AGVLGGPATP ASGPGPASAE PAVSEPGLGP DPKDTIIDVV GAPLTPNSRK
451 RRESASSSSS VKKAKKP
```

BLASTP hits

Entry ATAC2337_5 from database TREMBLNEW:
gene: "T08I13.5"; Arabidopsis thaliana chromosome II BAC T08I13 genomic sequence, complete sequence.
Score = 340, P = 2.6e-30, identities = 115/374, positives = 176/374

Entry YE8D_SCHPO from database SWISSPROT:
HYPOTHETICAL 47.1 KD PROTEIN C9G1.13C IN CHROMOSOME I.
Score = 221, P = 1.9e-20, identities = 67/192, positives = 97/192

Entry S64291 from database PIR:
hypothetical protein YGR002c - yeast (Saccharomyces cerevisiae)
Score = 202, P = 2.8e-13, identities = 71/260, positives = 124/260

Alert BLASTP hits for DKFZphfbr2_64c4, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_64c4, frame 2

Report for DKFZphfbr2_64c4.2

```
[LENGTH]        467
[MW]            53007.60
[pI]            9.51
[HOMOL]         TREMBL:ATAC2337_5 gene: "T08I13.5"; Arabidopsis thaliana chromosome II BAC
T08I13 genomic sequence, complete sequence. 4e-29
[FUNCAT]        99 unclassified proteins      [S. cerevisiae, YGR002c] 1e-19
[PROSITE]       MYRISTYL            1
[PROSITE]       CAMP_PHOSPHO_SITE   4
[PROSITE]       CK2_PHOSPHO_SITE    10
[PROSITE]       TYR_PHOSPHO_SITE    3
[PROSITE]       GLYCOSAMINOGLYCAN   1
[PROSITE]       PKC_PHOSPHO_SITE    12
[PROSITE]       ASN_GLYCOSYLATION   1
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY    20.13 %
```

```
SEQ    MATGADVRDILELGGPEGDAASGTISKKDIINPDKKKSKKSSETLTFKRPEGMHREVYAL
SEG    .......................xxxxxxxxxxxxxxxxx....................
PRD    ccceeeeeeeeeeccccccccccccccccccccccccccccccccccccccchhhhhhhh

SEQ    LYSDKKDAPPLLPSDTGQGYRTVKAKLGSKKVRPWKWMPFTNPARKDGAMFFHWRRAAEE
SEG    ............................................................
PRD    hhhhccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhc

SEQ    GKDYPFARFNKTVQEPVYSEQEYQLYLHDNAWTKAETDHLFDLSRRFDLRFVVIHDRYDH
SEG    ............................................................
PRD    cccccccccccccccccchhhhhhhhhhhcchhhhhhhhhhhhhhhhccceeeeeeecccc

SEQ    QQFKKRSVEDLKERYYHICAKLANVRAVPGTDLKIPVFDAGHERRRKEQLERLYNRTPEQ
SEG    ............................................................
PRD    chhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccchhhhhhhhhhhhhhhhhcchhh

SEQ    VAEEEYLLQELRKIEARKKEREKRSQDLQKLITAADTTAEQRRTERKAPKKKLPQKKEAE
SEG    .........xxxxxxxxxxxxxxx...................xxxxxxxxxxxxxxx
PRD    hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhh

SEQ    KPAVPETAGIKFPDFKSAGVTLRSQRMKLPSSVGQKKIKALEQMLLELGVELSPTPTEEL
SEG    xxx.........................................................
```

```
PRD      hccccccccccccccccceeehhhhhhhcccccchhhhhhhhhhhhhhhhccccccchhh

SEQ      VHMFNELRSDLVLLYELKQACANCEYELQMLRHRHEALARAGVLGGPATPASGPGPASAE
SEG      ............................................xxxxxxxxxxxxxxxx
PRD      hhhhhhcchhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhccccccccccccccccccc

SEQ      PAVSEPGLGPDPKDTIIDVVGAPLTPNSRKRRESASSSSSVKKAKKP
SEG      xxxxxxx....................xxxxxxxxxxxxxxxxxxx.
PRD      cccccccccccccceeeecccccccccccccccccccccceeecccc
```

Prosite for DKFZphfbr2_64c4.2

PS00001	130->134	ASN_GLYCOSYLATION	PDOC00001
PS00002	412->416	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	35->39	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	39->43	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	184->188	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	451->455	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00005	284->287	PKC_PHOSPHO_SITE	PDOC00005
PS00005	321->324	PKC_PHOSPHO_SITE	PDOC00005
PS00005	324->327	PKC_PHOSPHO_SITE	PDOC00005
PS00005	448->451	PKC_PHOSPHO_SITE	PDOC00005
PS00005	460->463	PKC_PHOSPHO_SITE	PDOC00005
PS00006	3->7	CK2_PHOSPHO_SITE	PDOC00006
PS00006	26->30	CK2_PHOSPHO_SITE	PDOC00006
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	153->157	CK2_PHOSPHO_SITE	PDOC00006
PS00006	187->191	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS00006	435->439	CK2_PHOSPHO_SITE	PDOC00006
PS00007	131->139	TYR_PHOSPHO_SITE	PDOC00007
PS00007	227->235	TYR_PHOSPHO_SITE	PDOC00007
PS00007	116->125	TYR_PHOSPHO_SITE	PDOC00007
PS00008	14->20	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_64c4.2)

DKFZphfbr2_64h6
----

group: brain derived

DKFZphfbr2_64h6 encodes a novel 176 amino acid protein with similarity to predicted yeast proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to S.pombe SPBC337.09 and S.cerevisiae YER044c

complete cDNA, complete cds accoring to YER044c/SPBC337.09, start at Bp 111, EST hits

Sequenced by Qiagen

Locus: /map="14"

Insert length: 1212 bp
Poly A stretch at pos. 1192, polyadenylation signal at pos. 1168

```
   1 GGGCTGGAGC TGTCCTGGGG GAGCTTGTTT GCGGCAGCGG CTGCTGCTGC
  51 CACTGCTGTG CTGGGGGCCC GGTCGCCAGG CAAAAAGCCC TCCCACGTTT
 101 GAGGGGAGTC ATGAGCCGTT TCCTGAATGT GTTAAGAAGT TGGCTGGTTA
 151 TGGTGTCCAT CATAGCCATG GGGAACACGC TGCAGAGCTT CCGAGACCAC
 201 ACTTTTCTCT ATGAAAAGCT CTACACTGGC AAGCCAAACC TTGTGAATGG
 251 CCTCCAAGCT CGGACCTTTG GGATCTGGAC GCTGCTCTCA TCAGTGATCC
 301 GCTGCCTCTG TGCCATTGAC ATTCACAACA AGACGCTCTA TCACATCACA
 351 CTCTGGACCT TCCTCCTTGC CCTGGGGCAT TTCCTCTCTG AGTTGTTTGT
 401 CTATGGAACT GCAGCTCCCA CGATTGGCGT CCTGGCACCC CTGATGGTGG
 451 CAAGTTTCTC CATCCTGGGT ATGCTGGTCG GGCTCCGGTA TCTAGAAGTA
 501 GAACCAGTAT CCAGACAGAA GAAGAGAAAC TGAGGCCAGC ATTATCACCT
 551 CCAGGACTTT CTCGTTTTCC ACCTTGGCCA TCTTCTTCCT TCGTCGTCTC
 601 TCCCCTTTAA TTTCTTTTCT ATTCCATCAT CTGCCCTTTT ACTCACTTTT
 651 AGCCTCTTTT TTTAATTTTT AAAATTTAAA GATATGCATA CTGAAAAGTA
 701 TATAACATGT ACGTACAATT TAAAGAATAA TTTTAAAGTG AATACTACGT
 751 AACTCCATCC AAGTCAAGAA ATTGCCAGCT TCTCGGAAGC CCACTGTGTC
 801 TCCTTCCCCT ACCTGCAACC TCTTCCAGGC TCCCTTTTCC AGCCTTCCCC
 851 TTTTTCCCTT TTATTTTCAT GCCTTGATTT GACTTGTGTG GTGGGAACAT
 901 GTGAACTATG AAACTTAAAC CTGCTGCCCA CCCAGAGCAG CTGTGACCAA
 951 GGGCTGCCTC AAGGGGTTGT CCACGCAGGT TGGGCTCCTC TCTGCTGCTG
1001 GACCCAAGAC TCTGAACCTT CCAAGGGACA GGCAGTTCTT CTGAGAAGGG
1051 CTCCCCTGTG TGTGAGCAAG ACCACAGCTC TCCTTCTATC TACAGATGCA
1101 TGAGGGTTGG AAGAGTCTGG GCTGTTTTTA GACCTTCTGG TCAGCTGTAT
1151 TTGTGTAACA ACTTTTGTAA TAAATAGAAA AACCCTCTGC TCAAAAAAAA
1201 AAAAAAAAAA AA
```

BLAST Results

Entry G38566 from database EMBL:
SHGC-64295 Human Homo sapiens STS genomic, sequence tagged site.
Score = 1398, P = 1.4e-56, identities = 284/288

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 0 bp to 530 bp; peptide length: 177
Category: similarity to unknown protein
Classification: unclassified

```
  1 AGAVLGELVC GSGCCCHCCA GGPVARQKAL PRLRGVMSRF LNVLRSWLVM
 51 VSIIAMGNTL QSFRDHTFLY EKLYTGKPNL VNGLQARTFG IWTLLSSVIR
101 CLCAIDIHNK TLYHITLWTF LLALGHFLSE LFVYGTAAPT IGVLAPLMVA
```

151 SFSILGMLVG LRYLEVEPVS RQKKRN

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64h6, frame 3

TREMBL:SPBC337_9 gene: "SPBC337.09"; product: "conserved hypothetical protein"; S.pombe chromosome II cosmid c337., N = 1, Score = 224, P = 1.4e-18

PIR:S50547 hypothetical protein YER044c - yeast (Saccharomyces cerevisiae), N = 1, Score = 192, P = 3.4e-15

```
>TREMBL:SPBC337_9 gene: "SPBC337.09"; product: "conserved hypothetical
     protein"; S.pombe chromosome II cosmid c337.
            Length = 136

  HSPs:

 Score = 224 (33.6 bits), Expect = 1.4e-18, P = 1.4e-18
 Identities = 49/113 (43%), Positives = 74/113 (65%)

Query:    42 NVLRSWLVMVSIIAMGNTLQSFRDHTFLYEKLYTGKPNLVNGLQARTFGIWTLLSSVIRC 101
             +++  W V+VS+ A+ NT+QSF     L +++Y+    N VNGLQ RTFGIWTLLS+++R
Sbjct:    11 SLVAKWNVVVSVAALFNTVQSFLTPK-LTKRVYSNT-NEVNGLQGRTFGIWTLLSAIVRF 68

Query:   102 LCAIDIHNKTLYHITLWTFLLALGHFLSELFVYGTAAPTIGVLAPLMVASFSI 154
              CA  I N  +Y +   T+ LA  HFLSE  ++ T     G+L+P++V++ SI
Sbjct:    69 YCAYHITNPDVYFLCQCTYYLACFHFLSEWLLFRTTNLGPGLLSPIVVSTVSI 121
```

Pedant information for DKFZphfbr2_64h6, frame 3

Report for DKFZphfbr2_64h6.3

```
[LENGTH]       176
[MW]           19359.31
[pI]           9.53
[HOMOL]        TREMBL:SPBC337_9 gene: "SPBC337.09"; product: "conserved hypothetical protein";
S.pombe chromosome II cosmid c337. 2e-17
[FUNCAT]       99 unclassified proteins     [S. cerevisiae, YER044c] 7e-16
[KW]           TRANSMEMBRANE  2
[KW]           LOW_COMPLEXITY    7.39 %

SEQ    AGAVLGELVCGSGCCCHCCAGGPVARQKALPRLRGVMSRFLNVLRSWLVMVSIIAMGNTL
SEG    .........xxxxxxxxxxxxx......................................
PRD    ccceeeeeeeecccceeeecccccccccccccccchhhhhhhhhhhhhhheeeecccccc
MEM    .......................................MMMMMMMMMMMMMMMMM....

SEQ    QSFRDHTFLYEKLYTGKPNLVNGLQARTFGIWTLLSSVIRCLCAIDIHNKTLYHITLWTF
SEG    ............................................................
PRD    cccccchhhhhhhhhhcccccccccccccccchhhhhhhhhhhhhhcccccceeeehhhhh
MEM    ............................................................

SEQ    LLALGHFLSELFVYGTAAPTIGVLAPLMVASFSILGMLVGLRYLEVEPVSRQKKRN
SEG    ........................................................
PRD    hhhhhhhhhhhhhhhcccccccccccceeehhhhhhhhhhhheeeeeccccccccccc
MEM    ......................MMMMMMMMMMMMMMMMM.................
```

(No Prosite data available for DKFZphfbr2_64h6.3)

(No Pfam data available for DKFZphfbr2_64h6.3)

DKFZphfbr2_64j18
----

group: Intracellular transport and trafficking

DKFZphfbr2_624j18.1 encodes a novel 180 amino acid protein nearly identical to the microsomal signal peptidase 23 kd subunit of canis familiaris, gallus gallus and C. elegans.

The new protein is identical to canine and chicken microsomal signal peptidase 23 kd subunit. The canine microsomal signal peptidase is a protein complex comprised of five subunits (25, 22/23, 21, 18, and 12 kDa). The 23kDa subunit is tightly associated with the 18- and 21-kDa subunits, that are integral membrane proteins.

The new protein can find application in modulation of protein transport into microsomal compartments and as a tool for proteomic analysis.

strong similarity to dog signal peptidase (EC 3.4.99.-)

complete cDNA, complete cds, potential start at Bp 109, EST hits,

Sequenced by Qiagen

Locus: unknown

Insert length: 690 bp
Poly A stretch at pos. 666, polyadenylation signal at pos. 646

```
  1 GCCGGAACGC GCGCACCGCA GACGGCGCGG ATCGCAGGGA GCCGGTCCGC
 51 CGCCGGAACG GGAGCCTGGG TGTGCGTGTG GAGTCCGGAC TCGTGGGAGA
101 CGATCGCGAT GAACACGGTG CTGTCGCGGG CGAACTCACT GTTCGCCTTC
151 TCGCTGAGCG TGATGGCGGC GCTCACCTTC GGCTGCTTCA TCACCACCGC
201 CTTCAAAGAC AGGAGCGTCC CGGTGCGGCT GCACGTCTCG CGGATCATGC
251 TAAAAAATGT AGAAGATTTC ACTGGACCTA GAGAAAGAAG TGATCTGGGA
301 TTTATCACAT CTGATATAAC TGCTGATCTA GAGAATATAT TTGATTGGAA
351 TGTTAAGCAG TTGTTTCTTT ATTTATCAGC AGAATATTCA ACAAAAAATA
401 ATGCTCTGAA CCAAGTTGTC CTATGGGACA AGATTGTTTT GAGAGGTGAT
451 AATCCGAAGC TGCTGCTGAA AGATATGAAA ACAAAATATT TTTTCTTTGA
501 CGATGGAAAT GGTCTCAAGG GAAACAGGAA TGTCACTTTG ACCCTGTCTT
551 GGAACGTCGT ACCAAATGCT GGAATTCTAC CTCTTGTGAC AGGATCAGGA
601 CACGTATCTG TCCCATTTCC AGATACATAT GAAATAACGA AGAGTTATTA
651 AATTATTCTG AATTTGAAAC AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

89034208:
cDNA-derived primary structure of the glycoprotein component of canine microsomal
signal peptidase complex.

Peptide information for frame 1

ORF from 109 bp to 648 bp; peptide length: 180
Category: strong similarity to known protein
Prosite motifs: TONB_DEPENDENT_REC_1 (1-58)
RGD (148-151)

```
  1 MNTVLSRANS LFAFSLSVMA ALTFGCFITT AFKDRSVPVR LHVSRIMLKN
 51 VEDFTGPRER SDLGFITSDI TADLENIFDW NVKQLFLYLS AEYSTKNNAL
101 NQVVLWDKIV LRGDNPKLLL KDMKTKYFFF DDGNGLKGNR NVTLTLSWNV
151 VPNAGILPLV TGSGHVSVPF PDTYEITKSY
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64j18, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_64j18, frame 1

Report for DKFZphfbr2_64j18.1

```
[LENGTH]        180
[MW]            20253.39
[pI]            8.66
[HOMOL]         PIR:A31788 signal peptidase (EC 3.4.99.-) (SPC 22/23) - dog 1e-100
[FUNCAT]        30.07 organization of endoplasmatic reticulum       [S. cerevisiae, YLR066w]
6e-15
[FUNCAT]        06.07 protein modification (glycolsylation, acylation, myristylation,
palmitylation, farnesylation and processing)        [S. cerevisiae, YLR066w] 6e-15
[PIRKW]         transmembrane protein 2e-92
[PIRKW]         glycoprotein 2e-92
[PIRKW]         hydrolase 2e-92
[PROSITE]       RGD     1
[PROSITE]       MYRISTYL        2
[PROSITE]       PROKAR_LIPOPROTEIN      1
[PROSITE]       TONB_DEPENDENT_REC_1    1
[PROSITE]       PKC_PHOSPHO_SITE        1
[PROSITE]       ASN_GLYCOSYLATION       1
[KW]            Alpha_Beta
[KW]            SIGNAL_PEPTIDE 32
```

```
SEQ   MNTVLSRANSLFAFSLSVMAALTFGCFITTAFKDRSVPVRLHVSRIMLKNVEDFTGPRER
PRD   cccccccchhhhhhhhhhhhhhhhhhhhhhheeecccccceeehhhhhhhhhhhhccccccc

SEQ   SDLGFITSDITADLENIFDWNVKQLFLYLSAEYSTKNNALNQVVLWDKIVLRGDNPKLLL
PRD   cccccchhhhhhhhccccccchhhhhhhhhhhhhhhcccccceeeeeeeceeeccccchhhhh

SEQ   KDMKTKYFFFDDGNGLKGNRNVTLTLSWNVVPNAGILPLVTGSGHVSVPFPDTYEITKSY
PRD   hhcccceeeeecccccccccceeeeeeeecccccceeeeecccccceeeecccccccccccc
```

Prosite for DKFZphfbr2_64j18.1

PS00001	141->145	ASN_GLYCOSYLATION	PDOC00001
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	135->141	MYRISTYL	PDOC00008
PS00013	16->27	PROKAR_LIPOPROTEIN	PDOC00013
PS00016	112->115	RGD	PDOC00016
PS00430	1->22	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKFZphfbr2_64j18.1)

DKFZphfbr2_64k24
----

group: transmembrane proteins

DKFZphfbr2_64k24 encodes a novel 412 amino acid protein with weak similarity to several known proteins.

The novel protein contains 5 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to AMAC1 "testicular condensing enzyme" ;
membrane regions: 5
Summary DKFZphfbr2_64k24 encodes a novel 412 amino acid protein, with similarity to AMAC1"; product: "testicular condensing enzyme

similarity to AMAC1 "testicular condensing enzyme"

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1958 bp
Poly A stretch at pos. 1939, polyadenylation signal at pos. 1918

```
   1 GGGCCCGCCT CGATTTTCCC AGGCGAGGGC ACGCCCGCGT CAGTCGCCTC
  51 CGGGGCACCT TCCTCGCCAC GACACGCAGG TAACCGGGCC CCGGGAGCCG
 101 GTCGGCGGCG GCGGACTGGG ACCTTGATCC TGCCTGCCCG GCCGCCCGAC
 151 AAGGGAATGA GAGCGGACCC CGAACTCCAC ACACCCGCGT TTAGCCGCCA
 201 CACCTAAGGG GCAGAACAGT CTTTTTGGGT AAGGGCCGGG CTGGGGGCGA
 251 CGCGCCCCGC CCGCTTTGCA GACTTCGGGG TGCTCTGCAC GACGCCTGAA
 301 AGGCCGCGGG GCCCGCATTT CTCTGTGCTG CCCTCCTGGA GAACCGGGAC
 351 ACGGGGACGG GAGGGCCAGC ATCGGCTACG GCCCGGTTTC CCGTTTCTTT
 401 CCTCTGTCGC GTCTGGGCCC TCCTGCAGCG TCCATGATGA AGGCCAGGGG
 451 CTGTTGCTTT CCTCTCGCCC AGTAGCCAAC CCAAGCAAGG GAATTAATTA
 501 TCTGAAGAAA TGGATACTTC TCCCTCCAGA AAATATCCAG TTAAAAAACG
 551 GGTGAAAATA CATCCCAACA CAGTGATGGT GAAATATACT TCTCATTATC
 601 CCCAGCCTGG CGATGATGGA TATGAAGAAA TCAATGAAGG CTATGGGAAT
 651 TTTATGGAGG AAAATCCAAA GAAAGGTCTG CTGAGTGAAA TGAAAAAAAA
 701 AGGGAGAGCT TTCTTTGGAA CCATGGATAC CCTACCTCCA CCAACAGAAG
 751 ACCCAATGAT CAATGAGATT GGACAATTCC AGAGCTTTGC AGAAAAAAAC
 801 ATTTTTCAAT CCCGAAAAAT GTGGATAGTG CTGTTTGGAT CTGCTTTGGC
 851 TCATGGATGT GTAGCTCTTA TCACTAGGCT TGTTTCTGAT CGGTCTAAAG
 901 TTCCATCTCT AGAACTGATT TTTATCCGTT CTGTTTTTCA GGTCTTATCT
 951 GTGTTAGTTG TGTGTTACTA TCAGGAGGCC CCCTTTGGAC CCAGTGGATA
1001 CAGATTACGA CTCTTCTTTT ATGGTGTATG CAATGTCATT TCTATCACTT
1051 GTGCTTATAC ATCATTTTCA ATAGTTCCTC CCAGCAATGG GACCACTATG
1101 TGGAGAGCCA CAACTACAGT CTTCAGTGCC ATTTTGGCTT TTTTACTCGT
1151 AGATGAGAAA ATGGCTTATG TTGACATGGC TACAGTTGTT TGCAGCATCT
1201 TAGGTGTTTG TCTTGTCATG ATCCCAAACA TTGTTGATGA AGACAATTCT
1251 TTGTTAAATG CCTGGAAAGA AGCCTTTGGG TACACCATGA CTGTGATGGC
1301 TGGACTGACC ACTGCTCTCT CAATGATAGT ATACAGATCC ATCAAGGAGA
1351 AGATCAGCAT GTGGACTGCG CTGTTTACTT TTGGTTGGAC TGGGACAATT
1401 TGGGGAATAT CTACTATGTT TATTCTTCAA GAACCCATCA TCCCATTAGA
1451 TGGAGAAACC TGGAGTTATC TCATTGCTAT ATGTGTCTGT TCTACTGCAG
1501 CATTCTTAGG AGTTTATTAT GCCTTGGACA AATTCCATCC AGCTTTGGTT
1551 AGCACAGTAC AACATTTGGA GATTGTGGTA GCTATGGTCT TGCAGCTTCT
1601 CGTGCTGCAC ATATTTCCTA GCATCTATGA TGTTTTTGGA GGGGTAATCA
1651 TTATGATTAG TGTTTTTGTC CTTGCTGGCT ATAAACTTTA CTGGAGGAAT
1701 TTAAGAAGGC AGGACTACCA GGAAATACTA GACTCTCCCA TTAAATGAAT
1751 ACCTGATTAT TATTGTCTCA TTAATGTTCA GTTATTAATA TGTATACTGC
1801 CATTTTAATG TTTACCTATG AATGTCTTTT GTGTTATATA ACTGACAGAG
1851 TGCTATAAAA TATATAATAT ATACAAATGC AGAAAATTTA TTCTAGTCTA
1901 ATATATTCAA ATACAAATAT TAAATATATG AAATACGTTA AAAAAAAAAA
1951 AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 510 bp to 1745 bp; peptide length: 412
Category: similarity to known protein

```
  1 MDTSPSRKYP VKKRVKIHPN TVMVKYTSHY PQPGDDGYEE INEGYGNFME
 51 ENPKKGLLSE MKKKGRAFFG TMDTLPPPTE DPMINEIGQF QSFAEKNIFQ
101 SRKMWIVLFG SALAHGCVAL ITRLVSDRSK VPSLELIFIR SVFQVLSVLV
151 VCYYQEAPFG PSGYRLRLFF YGVCNVISIT CAYTSFSIVP PSNGTTMWRA
201 TTTVFSAILA FLLVDEKMAY VDMATVVCSI LGVCLVMIPN IVDEDNSLLN
251 AWKEAFGYTM TVMAGLTTAL SMIVYRSIKE KISMWTALFT FGWTGTIWGI
301 STMFILQEPI IPLDGETWSY LIAICVCSTA AFLGVYYALD KFHPALVSTV
351 QHLEIVVAMV LQLLVLHIFP SIYDVFGGVI IMISVFVLAG YKLYWRNLRR
401 QDYQEILDSP IK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64k24, frame 3

TREMBLNEW:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus musculus testicular condensing enzyme (AMAC1) mRNA, complete cds., N = 1, Score = 191, P = 1.9e-12

TREMBL:BMAJ733_6 product: "hypothetical protein"; Bacillus megaterium bgaM gene, N = 1, Score = 137, P = 1.6e-06

PIR:G71841 hypothetical protein jhp1155 - Helicobacter pylori (strain J99), N = 1, Score = 129, P = 1.3e-05

>TREMBLNEW:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus musculus testicular condensing enzyme (AMAC1) mRNA, complete cds.
Length = 362

HSPs:

Score = 191 (28.7 bits), Expect = 1.9e-12, P = 1.9e-12
Identities = 39/105 (37%), Positives = 66/105 (62%)

```
Query:   289 FTFGWTGTIWGISTMFILQEPIIPLDGETWSYLIAICVCSTAAFLGVYYALDKFHPALVS 348
             F FG  G +  +  +F+LQ P++P D  +WS ++A+ + +  +F+ V YA+ K HPALV
Sbjct:   248 FLFGLVGLMVSVPGLFVLQTPVLPQDTLSWSCVVAVGLLALVSFVCVSYAVTKAHPALVC 307

Query:   349 TVQHLEIVVAMVLQLLVLH--IFPSIYDVFGGVIIMISVFVLAGYKL 393
              V H E+VVA++LQ  VL+  + PS  D+ G  +++ S+ ++    L
Sbjct:   308 AVLHSEVVVALMLQYYVLYETVAPS--DIMGAGVVLGSIAIITAQNL 352
```

Pedant information for DKFZphfbr2_64k24, frame 3

Report for DKFZphfbr2_64k24.3

```
[LENGTH]    412
[MW]        46449.87
[pI]        6.99
[HOMOL]     TREMBL:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus musculus testicular condensing enzyme (AMAC1) mRNA, complete cds. 8e-14
[PROSITE]   MYRISTYL    6
[PROSITE]   CK2_PHOSPHO_SITE    3
[PROSITE]   PKC_PHOSPHO_SITE    4
[PROSITE]   ASN_GLYCOSYLATION    1
[KW]        TRANSMEMBRANE  5

SEQ     MDTSPSRKYPVKKRVKIHPNTVMVKYTSHYPQPGDDGYEEINEGYGNFMEENPKKGLLSE
```

```
PRD  cccccccccccccceeeeccccceeeeeeccccccccccceeeeecccccccccccccccchhhh
MEM  ..............................................................

SEQ  MKKKGRAFFGTMDTLPPPTEDPMINEIGQFQSFAEKNIFQSRKMWIVLFGSALAHGCVAL
PRD  hhhhcceeeccccccccccccccceeeecccchhhhhhhhccceeeeeeeecccccchhhhhc
MEM  ..............................................................

SEQ  ITRLVSDRSKVPSLELIFIRSVFQVLSVLVVCYYQEAPFGPSGYRLRLFFYGVCNVISIT
PRD  chhhhhccccccccchhhhhhhhhhhheeeeeeeccccccccceeeeeeeeecceeeeeee
MEM  ................MMMMMMMMMMMMMMMMM..........................

SEQ  CAYTSFSIVPPSNGTTMWRATTTVFSAILAFLLVDEKMAYVDMATVVCSILGVCLVMIPN
PRD  eccceeeecccccccceeeeeehhhhhhhhhhhhhhhhheeeeeeeeeeeeeeeeeeeecc
MEM  ..............................................................

SEQ  IVDEDNSLLNAWKEAFGYTMTVMAGLTTALSMIVYRSIKEKISMWTALFTFGWTGTIWGI
PRD  ccccchhhhhhhhhhhheeeeeeehhhhhhcchhhhhhhhhhhccccccccccceeeec
MEM  .........................................MMMMMMMMMMMMMMMMMMM

SEQ  STMFILQEPIIPLDGETWSYLIAICVCSTAAFLGVYYALDKFHPALVSTVQHLEIVVAMV
PRD  ceeeeeecccccccccceeeeeccchhhhhhhhhccccccccccchhhhhhhhhhhhhh
MEM  MMMMMMMMMM.......MMMMMMMMMMMMMMMMMMMMMMMM......MMMMMMMMMMMMMM

SEQ  LQLLVLHIFPSIYDVFGGVIIMISVFVLAGYKLYWRNLRRQDYQEILDSPIK
PRD  hhhhhhhhcccccceeeeeeeeeeccccccchhhhhhhhhhhhhhhhccccc
MEM  MMMMMMM....MMMMMMMMMMMMMMMMMMMMMM....................
```

Prosite for DKFZphfbr2_64k24.3

PS00001	193->197	ASN_GLYCOSYLATION	PDOC00001
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS00005	126->129	PKC_PHOSPHO_SITE	PDOC00005
PS00005	277->280	PKC_PHOSPHO_SITE	PDOC00005
PS00006	92->96	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006
PS00008	70->76	MYRISTYL	PDOC00008
PS00008	88->94	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	265->271	MYRISTYL	PDOC00008
PS00008	295->301	MYRISTYL	PDOC00008
PS00008	334->340	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_64k24.3)

DKFZphfbr2_6a17

group: brain derived

DKFZphfbr2_6a17 encodes a novel 100 amino acid protein with very weak similarity to human finger protein zfOC1.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 1424 bp
Poly A stretch at pos. 1405, polyadenylation signal at pos. 1389

```
   1 GGGACTGAGG GGGTGGGCTT ACTCCCTGGG CAGTCTTGGG GGCCAGAGCT
  51 GAGGCCAGTC CATATTACAG TGGCTGGGCT GTTTTTTTCA GTAGCCCCTA
 101 GCATTGGCTG GGATTCCTGT TCCTGGGTGC GCCTCCACCT CCCTTCTGAT
 151 GCTTCCTGGC TATGGTGGGG TGGGAACCTC AGTTTCCCCC AAAGTCTTCC
 201 CTGGATGCTG GCTTCAGGTT GAAGACCCTG GTTCTTCCAG TTCCTCACGG
 251 GTTAGGTAGG GGCTCCTGCA TCACCTTCAG AATCAGTTCC AACCCCCACT
 301 CTCCTTAGGC TTTGTGCTCT GCTCTGCCCT GCCAGGCTGC CCTTGTCCAT
 351 GTGAGTAGCA TGGGCGGGTG GTGGGGACGG CAGTGGTGAT GAAGGGGGTG
 401 CACCACAGGC CTCATGAAGC AGTTCCCACA TGGGCGTGTG GCTGGGGCGT
 451 GGCCACCACA GAGCACATGG CTGTGTCTAG GCGCAAGCAC TTTAGCAGTA
 501 TCTGTTTACA TGCGCAAGGA TCAAGCCGAC TACCTGTGCT GTCTACTGGG
 551 ACAGCAGTCT CCGAGCTACT CCGTACCTCC CTCTGCCAGG TCGTGGAGTT
 601 AGGCCCCAGT CCCTACTTGT CACTGGTTCC CACTGTGCTC CTAACTGTGC
 651 AGCACCTGGG AGCTCTGGCC TGGGGCTGGA GGCCCTGGTA GGAGCTGCAG
 701 TTGGAGGCCG TTCTGTGCCC AGCAGCGGTG AGCGGCTCCC ATGGGCCCTG
 751 TGTCTGCAGG GAGCCAGGGC TGCGGCACAT GTGCTGTGAA ACTGGCACCC
 801 ACCTGGCGTG CTGCTGCCGC CACTTGCTTC CTGCAGCACC TCCTACCCTG
 851 CTCCGTGTCC TCCCTCTCCC CGCGCCTGGC TCAGGAGTGC TGGAAAAGCT
 901 CACGCCTCGG CCTGGGAGCC TGGCCTCTTG ATATACCTCG AGCTTCCCCT
 951 GTGCTCCCCA GCCCCAGGAC CACTGGCCCC TTGGCCTGAG GGGCTGGGGG
1001 CCCCACGACC TGCAGCGTCG AGTCCGGGAG AGAGCCCGGA GCGGCGTGCC
1051 ATCTCGGCTC GGCCTTGCTG AGAGCCTCCG CCCTGGCTTT CTCCCTGTCT
1101 GGTTTCAGTG GCTCACGTTG GTGCTACACA GCTAGAATAG ATATATTTAG
1151 AGAGAGAGAT ATTTTTAAGA CAAAGCCCAC AATTAGCTGT CCTTTAACAC
1201 CGCAGAACCC CCTCCCAGAA GAAGAGCGAT CCCTCGGACG GTCCGGGCGG
1251 GCACCCTCAG CCGGGCTCTT TGCAGAAGCA GCACCGCTGA CTGTGGGCCC
1301 GGCCCTCAGA TGTGTACATA TACGGCTATT TCCTATTTTA CTGTTCTTCA
1351 GATTTAGTAC TTGTAAATAA ACACACACAT TAAGGAGAGA TTAAACATTT
1401 TTGCCAAAAA AAAAAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 389 bp to 688 bp; peptide length: 100
Category: putative protein

```
 1 MKGVHHRPHE AVPTWACGWG VATTEHMAVS RRKHFSSICL HAQGSSRLPV
51 LSTGTAVSEL LRTSLCQVVE LGPSPYLSLV PTVLLTVQHL GALAWGWRPW
```

BLASTP hits

Entry S70007 from database PIR:
finger protein zfOC1 - human (fragment)
Length = 183
Score = 62 (21.8 bits), Expect = 0.24, Sum P(2) = 0.22
Identities = 18/47 (38%), Positives = 24/47 (51%)

Alert BLASTP hits for DKFZphfbr2_6a17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_6a17, frame 2

Report for DKFZphfbr2_6a17.2

[LENGTH]	100	
[MW]	10944.82	
[pI]	9.49	
[PROSITE]	MYRISTYL	2
[PROSITE]	PKC_PHOSPHO_SITE	2
[KW]	Alpha_Beta	

```
SEQ    MKGVHHRPHEAVPTWACGWGVATTEHMAVSRRKHFSSICLHAQGSSRLPVLSTGTAVSEL
PRD    cccccccccccccccccccccchhhhhhhhhcccccceeeccccccccceeeccccchhhhh

SEQ    LRTSLCQVVELGPSPYLSLVPTVLLTVQHLGALAWGWRPW
PRD    hhhhheeeeeccccccceeecchhhhhhhhhchhhhhcccc
```

Prosite for DKFZphfbr2_6a17.2

PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	45->48	PKC_PHOSPHO_SITE	PDOC00005
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	54->60	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_6a17.2)

DKFZphfbr2_6b24
----

group: metabolism

DKFZphfkd2_6b24 encodes a novel 334 amino acid protein with similarity to several bacterial dTDP-4-dehydrorhamnose reductases (EC 1.1.1.133).

The novel protein seems to be a human enzyme similar to dTDP-4-dehydrorhamnose reductases. EC 1.1.1.133 catalises the reaction: dTDP-6-deoxy-L-mannose + NADP(+) <=> dTDP-4-dehydro-6-deoxy-L-mannose + NADPH.

The new protein can find application in modulation of rhamnose metabolism and as a new enzyme for biotechnologic production processes.

similar to dTDP-6-deoxy-L-mannose-dehydrogenases

complete cDNA, EST hits, complete cds
Nucleotide sugars metabolism seems to be a dehydrogenase
localisation: region of primer A missing

Sequenced by AGOWA

Locus: /map="5"

Insert length: 2054 bp
Poly A stretch at pos. 2028, polyadenylation signal at pos. 2015

```
   1 GGGGGAGGCC CGCGTCGATC CTGGGTTGGA GGAGGTGGCG GCCGCTGAGG
  51 CTGCGGCGTG AAGACGGCGG GCATGGTGGG GCGGGAGAAA GAGCTCTCTA
 101 TACACTTTGT TCCCGGGAGC TGTCCGCTGG TGGAGGAGGA AGTTAACATC
 151 CCTAATAGGA GGGTTCTGGT TACTGGTGCC ACTGGGCTTC TTGGCAGAGC
 201 TGTACACAAA GAATTTCAGC AGAATAATTG GCATGCAGTT GGCTGTGGTT
 251 TCAGAAGAGC AAGACCAAAA TTTGAACAGG TTAATCTGTT GGATTCTAAT
 301 GCAGTTCATC ACATCATTCA TGATTTTCAG CCCCATGTTA TAGTACATTG
 351 TGCAGCAGAG AGAAGACCAG ATGTTGTAGA AAATCAGCCA GATGCTGCCT
 401 CTCAACTTAA TGTGGATGCT TCTGGGAATT TAGCAAAGGA AGCAGCTGCT
 451 GTTGGAGCAT TTCTCATCTA CATTAGCTCA GATTATGTAT TTGATGGAAC
 501 AAATCCACCT TACAGAGAGG AAGACATACC AGCTCCCCTA AATTTGTATG
 551 GCAAAACAAA ATTAGATGGA GAAAAGGCTG TCCTGGAGAA CAATCTAGGA
 601 GCTGCTGTTT TGAGGATTCC TATTCTGTAT GGGGAAGTTG AAAAGCTCGA
 651 AGAAAGTGCA GTGACTGTTA TGTTTGATAA AGTGCAGTTC AGCAACAAGT
 701 CAGCAAACAT GGATCACTGG CAGCAGAGGT TCCCCACACA TGTCAAAGAT
 751 GTGGCCACTG TGTGCCGGCA GCTAGCAGAG AAGAGAATGC TGGATCCATC
 801 AATTAAGGGA ACCTTTCACT GGTCTGGCAA TGAACAGATG ACTAAGTATG
 851 AAATGGCATG TGCAATTGCA GATGCCTTCA ACCTCCCCAG CAGTCACTTA
 901 AGACCTATTA CTGACAGCCC TGTCCTAGGA GCACAACGTC CGAGAAATGC
 951 TCAGCTTGAC TGCTCCAAAT TGGAGACCTT GGGCATTGGC CAACGAACAC
1001 CATTTCGAAT TGGAATCAAA GAATCACTTT GGCCTTTCCT CATTGACAAG
1051 AGATGGAGAC AAACGGTCTT TCATTAGTTT ATTTGTGTTG GGTTCTTTTT
1101 TTTTTTAAAT GAAAAGTATA GTATGTGGCC CTTTTTAAAG AACAAAGGAA
1151 ATAGTTTTGT ATGAGTACTT TAATTGTGAC TCTTAGGATC TTTCAGGTAA
1201 ATGATGCTCT TGCACTAGTG AAATTGTCTA AAGAAACTAA AGGGCAGTCA
1251 TGCCCTGTTT GCAGTAATTT TTCTTTTTAT CATTATGTTT GTCCTGGCTA
1301 AACTTGGAGT TTGAGTATAG TAAATTATGA TCCTTAAATA TTTGAGGGTC
1351 AGGATGAAGC AGATCTGCTG TAGACTTTTC AGATGAAATT GTTCATTCTC
1401 GTAACCTCCA TATTTCAGG ATTTTTGAAG CTGTTGACCA TTTCATGTTG
1451 ATTATTTTAA ATTGTGTGGA ATAGTATAAA AATCATTGGT GTTCATTATT
1501 TGCTTTGCCT GAGCTCAGAT CAAAATGTTT GAAGAAAGGA ACTTTATTTT
1551 TGCAAGTTAC GTACAGTTTT TATGCTTGAG ATATTTCAAC ATGTTATGTA
1601 TATTGGAACT TCTACAGCTT GATGCCTCCT GCTTTTATAG CAGTTTATGG
1651 GGAGCACTTG AAAGAGCGTG TGTACATGTA TTTTTTTTCT AGGCAAACAT
1701 TGAATGCAAA CGTGTATTTT TTTAATATAA ATATATAACT GTCCTTTTCA
1751 TCCCATGTTG CCGCTAAGTG ATATTTCATA TGTGTGGTTA TACTCATAAT
1801 AATGGGCCTT GTAAGTCTTT TCACCATTCA TGAATAATAA TAAATATGTA
1851 CTGCTGGCAT GTAATGCTTA GTTTTCTTGT ATTTACTTCT TTTTTTTAAA
1901 TGTAAGGACC AAACTTCTAA ACTAATTGTT CTTTTGTTGC TTTAATTTTT
1951 AAAAATTACA TTCTTCTGAT GTAACATGTG ATACATACAA AAGAATATAG
2001 TTTAATATGT ATTGAAATAA AACACAATAA AATTAAAAAA AAAAAAAAAA
2051 AAAA
```

BLAST Results

Entry G37115 from database EMBL:
SHGC-56899 Human Homo sapiens STS genomic.
Score = 446, P = 4.6e-14, identities = 90/91

Medline entries

99109950:
The metabolism of 6-deoxyhexoses in bacterial and animal cells.

Peptide information for frame 1

ORF from 73 bp to 1074 bp; peptide length: 334
Category: similarity to known protein

```
  1 MVGREKELSI HFVPGSCRLV EEEVNIPNRR VLVTGATGLL GRAVHKEFQQ
 51 NNWHAVGCGF RRARPKFEQV NLLDSNAVHH IIHDFQPHVI VHCAAERRPD
101 VVENQPDAAS QLNVDASGNL AKEAAAVGAF LIYISSDYVF DGTNPPYREE
151 DIPAPLNLYG KTKLDGEKAV LENNLGAAVL RIPILYGEVE KLEESAVTVM
201 FDKVQFSNKS ANMDHWQQRF PTHVKDVATV CRQLAEKRML DPSIKGTFHW
251 SGNEQMTKYE MACAIADAFN LPSSHLRPIT DSPVLGAQRP RNAQLDCSKL
301 ETLGIGQRTP FRIGIKESLW PFLIDKRWRQ TVFH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_6b24, frame 1

PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) - Actinobacillus actinomycetemcomitans, N = 1, Score = 293, P = 6.4e-26

TREMBL:SSU51197_21 gene: "rhsD"; product: "dTDP-6-deoxy-L-mannose-dehydrogenase"; Sphingomonas S88 sphingan polysaccharide synthesis (spsG), (spsS), (spsR), glycosyl transferase (spsQ), (spsI), glycosyl transferase (spsK), glycosyl transferase (spsL), (spsJ), (spsF), (spsD), (spsC), (spsE), Urf 32, Urf 26, ATP-binding cassette trans>., N = 1, Score = 291, P = 1e-25

SWISSPROT:RFBD_RHISN PROBABLE DTDP-4-DEHYDRORHAMNOSE REDUCTASE (EC 1.1.1.133) (DTDP-4-KETO- L-RHAMNOSE REDUCTASE) (DTDP-6-DEOXY-L-MANNOSE DEHYDROGENASE) (DTDP-L- RHAMNOSE SYNTHETASE)., N = 1, Score = 283, P = 7.4e-25

```
>PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) -
     Actinobacillus actinomycetemcomitans
          Length = 294

  HSPs:

 Score = 293 (44.0 bits), Expect = 6.4e-26, P = 6.4e-26
 Identities = 89/276 (32%), Positives = 151/276 (54%)

Query:    30 RVLVTGATGLLGRAVHKEFQQNNWHAVGCGFRRARPKFEQVNLLDSNAVHHIIHDFQPHV 89
             R+L+TGA G LGR++ K    N  + V          F ++++ + + V  II  F+P+V
Sbjct:     3 RLLITGAGGQLGRSLAKLLVDNGRYEV------LALDFSELDITNKDMVFSIIDSFKPNV 56

Query:    90 IVHCAAERRPDVVENQPDAASQLNVDASGNLAKEAAAVGAFLIYISSDYVFDG-TNPPYR 148
             I++ AA    D  E +  +A  +NV     LA+ A    + ++++S+DYVFDG  +  Y+
Sbjct:    57 IINAAAYTSVDQAELEVSSAYSVNVRGVQYLAEAAIRHNSAILHVSTDYVFDGYKSGKYK 116

Query:   149 EEDIPAPLNLYGKTKLDGEKAVLENNLGAAVLRIPILYGEVEKLEESAVTVMFDKVQFSN 208
             E DI  PL +YGK+K +GE+ +L  +  + +LR    +GE      + V  M  ++  +
Sbjct:   117 ETDIIHPLCVYGKSKAEGERLLLTLSPKSIILRTSWTFGEYGN---NFVKTML-RLAKNR 172

Query:   209 KSANMDHWQQRFPTHVKDVATVCRQLAEKRMLDPSIK-GTFHWSGNEQMTKYEMACAIAD 267
                 +   Q   PT+  D+A+V  Q+AEK ++  ++K G +H++G   ++ Y+ A AI D
Sbjct:   173 DILGVVADQIGGPTYSGDIASVLIQIAEKIIVGETVKYGIYHFTGEPCVSWYDFAIAIFD 232

Query:   268 AF-------NLPSSHLRPITDSPVLGAQRPRNAQLDCSKLE-TLGI 305
                      N+P  +     D P L A+RP N+ LD +K++   GI
Sbjct:   233 EAVAQKVLENVPLVNAITTADYPTL-AKRPANSCLDLTKIQQAFGI 277
```

Pedant information for DKFZphfbr2_6b24, frame 1

Report for DKFZphfbr2_6b24.1

```
[LENGTH]        334
[MW]            37551.98
[pI]            6.90
[HOMOL]         PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) -
Actinobacillus actinomycetemcomitans 6e-25
[FUNCAT]        01.06.01 lipid, fatty-acid and sterol biosynthesis  [S. cerevisiae, YGL001c]
6e-04
[EC]            1.1.1.133 dTDP-4-dehydrorhamnose reductase 2e-16
[PIRKW]         lipopolysaccharide biosynthesis 2e-16
[PIRKW]         NADP 2e-16
[PIRKW]         oxidoreductase 2e-16
[PIRKW]         streptomycin biosynthesis 1e-19
[SUPFAM]        dTDP-dihydrostreptose synthase 1e-20
[PROSITE]       MYRISTYL        1
[PROSITE]       CK2_PHOSPHO_SITE        4
[PROSITE]       PKC_PHOSPHO_SITE        3
[PROSITE]       ASN_GLYCOSYLATION       1
[KW]            Alpha_Beta
```

```
SEQ     MVGREKELSIHFVPGSCRLVEEEVNIPNRRVLVTGATGLLGRAVHKEFQQNNWHAVGCGF
PRD     ccccccceeeccccccceeeeecccccccceeeeeccccchhhhhhhhhhccceeeeecc

SEQ     RRARPKFEQVNLLDSNAVHHIIHDFQPHVIVHCAAERRPDVVENQPDAASQLNVDASGNL
PRD     ccccccccccccchhhhhhhhhhccceeeehhhhhhhhhhhhhhhhhhhhhhcccchhh

SEQ     AKEAAAVGAFLIYISSDYVFDGTNPPYREEDIPAPLNLYGKTKLDGEKAVLENNLGAAVL
PRD     hhhhhhhhheeeeeeccccccccccccccccccccccchhhhhhhhcccccccceeee

SEQ     RIPILYGEVEKLEESAVTVMFDKVQFSNKSANMDHWQQRFPTHVKDVATVCRQLAEKRML
PRD     eeeeeecccccccchhhhhhhhhhhhccceeeccccccccccchhhhhhhhhhhhhh

SEQ     DPSIKGTFHWSGNEQMTKYEMACAIADAFNLPSSHLRPITDSPVLGAQRPRNAQLDCSKL
PRD     ccccccceeeecccccchhhhhhhhhhhhcccccccccccccccccccccchhhhhh

SEQ     ETLGIGQRTPFRIGIKESLWPFLIDKRWRQTVFH
PRD     hhhhcccccchhhhhhhhhhhhhhhhhhhhcccc
```

Prosite for DKFZphfbr2_6b24.1

PS00001	208->212	ASN_GLYCOSYLATION	PDOC00001
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	207->210	PKC_PHOSPHO_SITE	PDOC00005
PS00005	243->246	PKC_PHOSPHO_SITE	PDOC00005
PS00006	162->166	CK2_PHOSPHO_SITE	PDOC00006
PS00006	251->255	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	298->302	CK2_PHOSPHO_SITE	PDOC00006
PS00008	314->320	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_6b24.1)

DKFZphfbr2_6i20
----

group: brain derived

DKFZphfbr2_6i20 encodes a novel 296 amino acid protein with similarity to ribosomal protein L15 precursor of S. cerevisiae mitochondria.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to ribosomal protein L15 precursor, mitochondrial

complete cDNA, complete cds, EST hits
potential miochondrial L15 ribosomal protein

Sequenced by AGOWA

Locus: /map="377.5 cR from top of Chr8 linkage group"

Insert length: 1122 bp
Poly A stretch at pos. 1099, polyadenylation signal at pos. 1071

```
   1 GGGGGCCCTT GAAAGTTCTT GGATCTGCGG GTTATGGCCG GTCCCTTGCA
  51 GGGCGGTGGG GCCCGGGCCC TGGACCTACT CCGGGGCCTG CCGCGTGTGA
 101 GCCTGGCCAA CTTAAAGCCG AATCCCGGCT CCAAGAAACC GGAGAGAAGA
 151 CCAAGAGGTC GGAGAAGAGG TAGAAAATGT GGCAGAGGCC ATAAAGGAGA
 201 AAGGCAAAGA GGAACCCGGC CCCGCTTGGG CTTTGAGGGA GGCCAGACTC
 251 CATTTTACAT CCGAATCCCA AAATACGGGT TTAACGAAGG ACATAGTTTC
 301 AGACGCCAGT ATAAGCCTAT GAGTCTCAAT AGACTGCAGT ATCTTATTGA
 351 TTTGGGTCGT GTTGATCCTA GTCAACCTAT TGACTTAACC CAGCTTGTCA
 401 ATGGGAGAGG TGTGACCATC CAGCCACTTA AAAGGGATTA TGATGTCCAG
 451 CTGGTTGAGG AGGGTGCTGA CACCTTTACG GCAAAAGTTA ATATTGAAGT
 501 ACAGTTGGCT TCAGAACTAG CTATTGCTGC CATTGAAAAA AATGGTGGTG
 551 TTGTTACTAC AGCCTTCTAT GATCCAAGAA GTCTGGACAT TGTATGCAAA
 601 CCTGTTCCAT TCTTTCTTCG TGGACAACCC ATTCCAAAAA GAATGCTTCC
 651 ACCAGAAGAA CTGGTACCAT ATTACACTGA TGCAAAGAAC CGTGGGTACC
 701 TGGCGGATCC TGCCAAATTT CCTGAAGCAC GACTTGAACT CGCCAGGAAG
 751 TATGGTTATA TCTTACCTGA TATCACTAAA GATGAACTCT TCAAAATGCT
 801 CTGTACTAGG AAGGATCCAA GGCAGATTTT CTTTGGTCTT GCTCCAGGAT
 851 GGGTGGTGAA TATGGCCGAT AAGAAAATCC TAAAACCTAC AGATGAAAAT
 901 CTCCTTAAGT ATTATACCTC ATGAATTCCC GTCCAAGGAA GCAGAGTTGT
 951 TAAAGAGTAC TGGAATAGGG GCTGAAGGAT CTATATTCCC TTATTGCATT
1001 TTCCTTATGT ATAATTTTCC AGATGGTGAT GTTACTTTTC AGTGTACTCA
1051 TATGTCTCAT TTTCATCTAA AATTAAATGG CAGGAAACAA GGACTGCATA
1101 GAGAAAAAAA AAAAAAAAAA AA
```

BLAST Results

Entry HS500354 from database EMBL:
human STS WI-12392.
Length = 426
Minus Strand HSPs:
Score = 1791 (268.7 bits), Expect = 1.1e-74, P = 1.1e-74
Identities = 375/384 (97%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 34 bp to 921 bp; peptide length: 296
Category: strong similarity to known protein

```
1 MAGPLQGGGA RALDLLRGLP RVSLANLKPN PGSKKPERRP RGRRRGRKCG
```

```
 51 RGHKGERQRG TRPRLGFEGG QTPFYIRIPK YGFNEGHSFR RQYKPMSLNR
101 LQYLIDLGRV DPSQPIDLTQ LVNGRGVTIQ PLKRDYDVQL VEEGADTFTA
151 KVNIEVQLAS ELAIAAIEKN GGVVTTAFYD PRSLDIVCKP VPFFLRGQPI
201 PKRMLPPEEL VPYYTDAKNR GYLADPAKFP EARLELARKY GYILPDITKD
251 ELFKMLCTRK DPRQIFFGLA PGWVVNMADK KILKPTDENL LKYYTS
```

BLASTP hits

Entry S63258 from database PIR:
ribosomal protein L15 precursor, mitochondrial - yeast (Saccharomyces cerevisiae)
Length = 322
Score = 259 (91.2 bits), Expect = 2.0e-22, P = 2.0e-22
Identities = 71/200 (35%), Positives = 106/200 (53%)

Entry H70161 from database PIR:
ribosomal protein L15 (rplO) - Lyme disease spirochete
Length = 145
Score = 173 (60.9 bits), Expect = 4.8e-13, P = 4.8e-13
Identities = 45/140 (32%), Positives = 73/140 (52%)

Alert BLASTP hits for DKFZphfbr2_6i20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_6i20, frame 1

Report for DKFZphfbr2_6i20.1

```
[LENGTH]     296
[MW]         33495.98
[pI]         9.98
[HOMOL]      TREMBL:AF067212_1 gene: "F37F2.1"; Caenorhabditis elegans cosmid F37F2. 1e-38
[FUNCAT]     05.01 ribosomal proteins    [S. cerevisiae, YNL284c] 7e-15
[FUNCAT]     30.16 mitochondrial organization    [S. cerevisiae, YNL284c] 7e-15
[FUNCAT]     j mrna translation and ribosome biogenesis   [M. genitalium, MG169] 1e-06
[BLOCKS]     BL00475D
[BLOCKS]     BL00475B Ribosomal protein L15 proteins
[PIRKW]      ribosome 2e-13
[PIRKW]      mitochondrion 2e-13
[PIRKW]      protein biosynthesis 2e-13
[SUPFAM]     Escherichia coli ribosomal protein L15 4e-06
[PROSITE]    MYRISTYL     3
[PROSITE]    AMIDATION    2
[PROSITE]    CK2_PHOSPHO_SITE    2
[PROSITE]    PKC_PHOSPHO_SITE    4
[KW]         Alpha_Beta
[KW]         LOW_COMPLEXITY   12.50 %

SEQ    MAGPLQGGGARALDLLRGLPRVSLANLKPNPGSKKPERRPRGRRRGRKCGRGHKGERQRG
SEG    ......xxxxxxxxxxxxx..............xxxxxxxxxxxxxxxxxxxxxxxx...
PRD    cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ    TRPRLGFEGGQTPFYIRIPKYGFNEGHSFRRQYKPMSLNRLQYLIDLGRVDPSQPIDLTQ
SEG    ............................................................
PRD    cccccccccccccceeeeccccccccccccccccccchhhhhhhhcccccccccccccee

SEQ    LVNGRGVTIQPLKRDYDVQLVEEGADTFTAKVNIEVQLASELAIAAIEKNGGVVTTAFYD
SEG    ............................................................
PRD    ecccceeeecccccccceeeeeecccccccchhhhhhhhhhhhhhhhhhccceeeeeecc

SEQ    PRSLDIVCKPVPFFLRGQPIPKRMLPPEELVPYYTDAKNRGYLADPAKFPEARLELARKY
SEG    ............................................................
PRD    cccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhh

SEQ    GYILPDITKDELFKMLCTRKDPRQIFFGLAPGWVVNMADKKILKPTDENLLKYYTS
SEG    ........................................................
PRD    ccccccchhhhhhhhhcccccceeeeecccccceeeecccceeecccchhhhhcccc
```

Prosite for DKFZphfbr2_6i20.1

```
PS00005    33->36    PKC_PHOSPHO_SITE    PDOC00005
PS00005    88->91    PKC_PHOSPHO_SITE    PDOC00005
```

PS00005	149->152	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00006	248->252	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	171->177	MYRISTYL	PDOC00008
PS00008	268->274	MYRISTYL	PDOC00008
PS00009	41->45	AMIDATION	PDOC00009
PS00009	45->49	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_6i20.1)

DKFZphfbr2_6o17

group: nucleic acid management

DKFZphfbr2_6o17 encodes a novel 455 amino acid protein with strong similarity to DEAD-box ATP-dependent RNA helicases YHR065c and T26G10.1.

The S. cerevisiae protein YHR065c is required for maturation of the 35S RNA primary transcript.

The new protein can find application in modulating rRNA maturation.

strong similar to RNA helicases

complete cDNA, complete cds, EST hits
probable start at Bp 27 matchs kozak consensus ANNatgG
involved in maturation of r-RNA ??
YHR065c/Rrp3p is involved in maturation of the 35S primary transcript
Drs1p cold-sensitive mutation has slow 27S to 25S pre-rRNA
conversion and is deficient in 60S ribosomal subunits

Sequenced by AGOWA

Locus: unknown

Insert length: 1840 bp
Poly A stretch at pos. 1815, polyadenylation signal at pos. 1793

```
   1 GGGGACTTCC GGAGACCTCA CACAAGATGG CGGCACCCGA GGAACACGAT
  51 TCTCCGACCG AAGCGTCCCA GCCGATTGTG GAAGAGGAGG AAACTAAAAC
 101 ATTTAAAGAC CTGGGTGTGA CAGATGTGTT GTGTGAAGCT TGTGACCAGT
 151 TGGGATGGAC AAAACCCACC AAGATTCAGA TTGAAGCTAT TCCTTTGGCC
 201 TTACAAGGTC GTGATATCAT TGGGCTTGCA GAAACTGGCT CTGGAAAGAC
 251 AGGCGCCTTT GCTTTGCCCA TTCTAAACGC ACTGCTGGAG ACCCCGCAGC
 301 GTTTGTTTGC CCTAGTTCTT ACCCCGACTC GGGAGCTGGC CTTTCAGATC
 351 TCAGAGCAGT TTGAAGCCCT GGGGTCCTCT ATTGGAGTGC AGAGTGCTGT
 401 GATTGTAGGT GGAATTGATT CAATGTCTCA ATCTTTGGCC CTTGCAAAAA
 451 AACCACATAT AATAATAGCA ACTCCTGGTC GACTGATTGA CCACTTGGAA
 501 AATACGAAAG GTTTCAACTT GAGAGCTCTC AAATACTTGG TCATGGATGA
 551 AGCCGACCGA ATACTGAATA TGGATTTTGA GACAGAGGTT GACAAGATCC
 601 TCAAAGTGAT TCCTCGAGAT CGGAAAACAT TCCTCTTCTC TGCCACCATG
 651 ACCAAGAAGG TTCAAAAACT TCAGCGAGCA GCTCTGAAGA ATCCTGTGAA
 701 ATGTGCCGTT TCCTCTAAAT ACCAGACAGT TGAAAAATTA CAGCAATATT
 751 ATATTTTTAT TCCCTCTAAA TTCAAGGATA CCTACCTGGT TTATATTCTA
 801 AATGAATTGG CTGGAAACTC CTTTATGATA TTCTGCAGCA CCTGTAATAA
 851 TACCCAGAGA ACAGCTTTGC TACTGCGAAA TCTTGGCTTC ACTGCCATCC
 901 CCCTCCATGG ACAAATGAGT CAGAGTAAGC GCCTAGGATC CCTTAATAAG
 951 TTTAAGGCCA AGGCCCGTTC CATTCTTCTA GCAACTGACG TTGCCAGCCG
1001 AGGTTTGGAC ATACCTCATG TAGATGTGGT TGTCAACTTT GACATTCCTA
1051 CCCATTCCAA GGATTACATC CATCGAGTAG GTCGAACAGC TAGAGCTGGG
1101 CGCTCCGGAA AGGCTATTAC TTTTGTCACA CAGTATGATG TGGAACTCTT
1151 CCAGCGCATA GAACACTTAA TTGGGAAGAA ACTACCAGGT TTTCCAACAC
1201 AGGATGATGA GGTTATGATG CTGACAGAAC GCGTCGCTGA AGCCCAAAGG
1251 TTTGCCCGAA TGGAGTTAAG GGAGCATGGA GAAAAGAAGA AACGCTCGCG
1301 AGAGGATGCT GGAGATAATG ATGACACAGA GGGTGCTATT GGTGTCAGGA
1351 ACAAGGTGGC TGGAGGAAAA ATGAAGAAGC GGAAAGGCCG TTAATCACTT
1401 TTATGAAGGC TCGAGTTCTG CTGTTCTGTA AAAGAAAATT GGAGAATGAA
1451 ACCTGCTCCA ACAGAGATCA TGAGACTGAA ATTGGTCAGA ATTGTGTCCA
1501 GAATGTGCTC AGCTAATTCA GTATTCTTCC CCATTCTGGG TTGGAGTTTA
1551 CTGCAGAGTA ATTCTTACAG TGCTGATGTC AAGACTGTTA CTGTTCTTCG
1601 ACTTTGATTC CTTGCTCATG ACATGAGTAG GGTGTGCTCT TCTGTCACTT
1651 CACACAGACC TTTTGCCTTT TTTAGCTGCA AGTCAAGGAC TAGGTTGATG
1701 ATGCCCATGA CCTGTAATTG TAAAGAAGCT TGGACATCTG CAAATGATAT
1751 TTAAACCATC TTGGCTTGTG CTTTATTCAA ACTAATGTGA AACAATAAAT
1801 TTAAATATTA TTTTTAAAAG AAAAAAAAAA AAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 27 bp to 1391 bp; peptide length: 455
Category: strong similarity to known protein

```
  1 MAAPEEHDSP TEASQPIVEE EETKTFKDLG VTDVLCEACD QLGWTKPTKI
 51 QIEAIPLALQ GRDIIGLAET GSGKTGAFAL PILNALLETP QRLFALVLTP
101 TRELAFQISE QFEALGSSIG VQSAVIVGGI DSMSQSLALA KKPHIIIATP
151 GRLIDHLENT KGFNLRALKY LVMDEADRIL NMDFETEVDK ILKVIPRDRK
201 TFLFSATMTK KVQKLQRAAL KNPVKCAVSS KYQTVEKLQQ YYIFIPSKFK
251 DTYLVYILNE LAGNSFMIFC STCNNTQRTA LLLRNLGFTA IPLHGQMSQS
301 KRLGSLNKFK AKARSILLAT DVASRGLDIP HVDVVVNFDI PTHSKDYIHR
351 VGRTARAGRS GKAITFVTQY DVELFQRIEH LIGKKLPGFP TQDDEVMMLT
401 ERVAEAQRFA RMELREHGEK KKRSREDAGD NDDTEGAIGV RNKVAGGKMK
451 KRKGR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_6o17, frame 3

PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - Caenorhabditis elegans, N = 1, Score = 1497, P = 1.6e-153

PIR:S46713 hypothetical protein YHR065c - yeast (Saccharomyces cerevisiae), N = 1, Score = 1154, P = 3.6e-117

TREMBL:ATH010462_1 gene: "RH10"; product: "RNA helicase"; Arabidopsis thaliana mRNA for DEAD box RNA helicase, RH10, N = 1, Score = 1122, P = 8.9e-114

TREMBL:AC002985_2 product: "R27090_2"; Human DNA from chromosome 19-specific cosmid R27090, genomic sequence, complete sequence., N = 1, Score = 950, P = 1.5e-95

```
>PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - Caenorhabditis
     elegans
          Length = 489

  HSPs:

 Score = 1497 (224.6 bits), Expect = 1.6e-153, P = 1.6e-153
 Identities = 283/442 (64%), Positives = 364/442 (82%)

Query:    19 EEEETKTFKDLGVTDVLCEACDQLGWTKPTKIQIEAIPLALQGRDIIGLAETGSGKTGAF 78
             E+ + K+F +LGV+  LC+AC +LGW KP+KIQ  A+P ALQG+D+IGLAETGSGKTGAF
Sbjct:    39 EDVKEKSFAELGVSQPLCDACQRLGWMKPSKIQQAALPHALQGKDVIGLAETGSGKTGAF 98

Query:    79 ALPILNALLETPQRLFALVLTPTRELAFQISEQFEALGSSIGVQSAVIVGGIDSMSQSLA 138
             A+P+L +LL+ PQ  F LVLTPTRELAFQI +QFEALGS IG+ +AVIVGG+D  +Q++A
Sbjct:    99 AIPVLQSLLDHPQAFFCLVLTPTRELAFQIGQQFEALGSGIGLIAAVIVGGVDMAAQAMA 158

Query:   139 LAKKPHIIIATPGRLIDHLENTKGFNLRALKYLVMDEADRILNMDFETEVDKILKVIPRD 198
             LA++PHII+ATPGRL+DHLENTKGFNL+ALK+L+MDEADRILNMDFE E+DKILKVIPR+
Sbjct:   159 LARRPHIIVATPGRLVDHLENTKGFNLKALKFLIMDEADRILNMDFEVELDKILKVIPRE 218

Query:   199 RKTFLFSATMTKKVQKLQRAALKNPVKCAVSSKYQTVEKLQQYYIFIPSKFKDTYLVYIL 258
             R+T+LFSATMTKKV KL+RA+L++P + +VSS+Y+TV+ L+Q+YIF+P+K+K+TYLVY+L
Sbjct:   219 RRTYLFSATMTKKVSKLERASLRDPARVSVSSRYKTVDNLKQHYIFVPNKYKETYLVYLL 278

Query:   259 NELAGNSFMIFCSTCNNTQRTALLLRNLGFTAIPLHGQMSQSKRLGSLNKFKAKARSILL 318
             NE AGNS ++FC+TC  T + A++LR LG  A+PLHGQMSQ KRLGSLNKFK+KAR IL+
Sbjct:   279 NEHAGNSAIVFCATCATTMQIAVMLRQLGMQAVPLHGQMSQEKRLGSLNKFKSKAREILV 338

Query:   319 ATDVASRGLDIPHVDVVVNFDIPTHSKDYIHRVGRTARAGRSGKAITFVTQYDVELFQRI 378
              TDVA+RGLDIPHVD+V+N+D+P+ SKDY+HRVGRTARAGRSG AIT VTQYDVE +Q+I
Sbjct:   339 CTDVAARGLDIPHVDMVINYDMPSQSKDYVHRVGRTARAGRSGIAITVVTQYDVEAYQKI 398

Query:   379 EHLIGKKLPGFPTQDDEVMMLTERVAEAQRFARMELREHGEKKK-----RSREDAGDNDD 433
             E  +GKKL  +   ++EVM+L ER EA   AR+E++E  EKKK      R  +D GD ++
Sbjct:   399 EANLGKKLDEYKCVENEVMVLVERTQEATENARIEMKEMDEKKKSGKKRRQNDDFGDTEE 458

Query:   434 TEGAIGVRNKVAGGKMKKRKGR 455
```

```
              + G    +  K   GG+      GR
Sbjct:    459 SGGRFKMGIKSMGGRGGSGGGR 480
```

Pedant information for DKFZphfbr2_6o17, frame 3

Report for DKFZphfbr2_6o17.3

```
[LENGTH]     455
[MW]         50646.80
[pI]         9.18
[HOMOL]      PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - Caenorhabditis elegans
1e-167
[FUNCAT]     04.01.04 rrna processing      [S. cerevisiae, YHR065c] 1e-127
[FUNCAT]     30.10 nuclear organization    [S. cerevisiae, YHR065c] 1e-127
[FUNCAT]     04.99 other transcription activities [S. cerevisiae, YHR169w] 2e-79
[FUNCAT]     06.10 assembly of protein complexes   [S. cerevisiae, YLL008w] 1e-71
[FUNCAT]     04.05.01.07 chromatin modification    [S. cerevisiae, YMR290c] 4e-66
[FUNCAT]     j mrna translation and ribosome biogenesis   [H. influenzae, HI0231 RNA] 1e-63
[FUNCAT]     09.01 biogenesis of cell wall         [S. cerevisiae, YJL033w] 1e-58
[FUNCAT]     04.05.03 mrna processing (splicing)   [S. cerevisiae, YDL084w] 1e-55
[FUNCAT]     05.04 translation (initiation, elongation and termination)  [S. cerevisiae,
YOR204w] 5e-55
[FUNCAT]     30.03 organization of cytoplasm       [S. cerevisiae, YOR204w] 5e-55
[FUNCAT]     l genome replication, transcription, recombination and repair     [H.
influenzae, HI0892] 9e-48
[FUNCAT]     98 classification not yet clear-cut  [S. cerevisiae, YLR276c] 2e-45
[FUNCAT]     30.16 mitochondrial organization      [S. cerevisiae, YDR194c] 4e-42
[FUNCAT]     99 unclassified proteins      [S. cerevisiae, YGL064c] 7e-16
[FUNCAT]     03.19 recombination and dna repair   [S. cerevisiae, YMR190c] 7e-12
[FUNCAT]     11.10 cell death       [S. cerevisiae, YMR190c] 7e-12
[FUNCAT]     r general function prediction        [M. jannaschii, MJ1401] 5e-06
[BLOCKS]     BL00175B Phosphoglycerate mutase family phosphohistidine proteins
[BLOCKS]     BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]     BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]     BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]     BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW]      nucleus 4e-60
[PIRKW]      RNA binding 7e-69
[PIRKW]      DEAD box 7e-69
[PIRKW]      transmembrane protein 9e-41
[PIRKW]      DNA binding 3e-55
[PIRKW]      recF recombination pathway 3e-11
[PIRKW]      ATP 1e-126
[PIRKW]      purine nucleotide binding 7e-69
[PIRKW]      P-loop 1e-126
[PIRKW]      hydrolase 1e-55
[PIRKW]      protein biosynthesis 7e-69
[PIRKW]      ATP binding 3e-61
[SUPFAM]     ATP-dependent RNA helicase eIF-4A 8e-06
[SUPFAM]     WW repeat homology 4e-58
[SUPFAM]     translation initiation factor eIF-4A 7e-69
[SUPFAM]     DEAD/H box helicase homology 1e-126
[SUPFAM]     recQ helicase homology 5e-12
[SUPFAM]     ATP-dependent RNA helicase homology 8e-06
[SUPFAM]     unassigned DEAD/H box helicases 1e-126
[SUPFAM]     ATP-dependent RNA helicase DBP1 4e-60
[SUPFAM]     ATP-dependent RNA helicase DHH1 1e-58
[SUPFAM]     recQ protein 3e-11
[SUPFAM]     tobacco ATP-dependent RNA helicase DB10 4e-58
[SUPFAM]     Bloom's syndrome helicase 5e-12
[PROSITE]    DEAD_ATP_HELICASE       1
[PROSITE]    ATP_GTP_A       1
[PROSITE]    MYRISTYL        5
[PROSITE]    AMIDATION       1
[PROSITE]    CAMP_PHOSPHO_SITE       1
[PROSITE]    CK2_PHOSPHO_SITE        6
[PROSITE]    PKC_PHOSPHO_SITE        9
[PROSITE]    ASN_GLYCOSYLATION       1
[PFAM]       Helicases conserved C-terminal domain
[PFAM]       DEAD and DEAH box helicases
[KW]         Alpha_Beta
```

```
SEQ    MAAPEEHDSPTEASQPIVEEEETKTFKDLGVTDVLCEACDQLGWTKPTKIQIEAIPLALQ
PRD    cccccccccccccccchhhhhhhhhhhcccccchhhhhhhhhhcccccccccccccccccc

SEQ    GRDIIGLAETGSGKTGAFALPILNALLETPQRLFALVLTPTRELAFQISEQFEALGSSIG
PRD    ccceeeeeccccccceeehhhhhhhhcccccceeeeeecchhhhhhhhhhhhhhhhhhcc
```

```
SEQ   VQSAVIVGGIDSMSQSLALAKKPHIIIATPGRLIDHLENTKGFNLRALKYLVMDEADRIL
PRD   eeeeeeecchhhhhhhhhhccceeeeeecccccccccccccccccccceeehhhhhhhh

SEQ   NMDFETEVDKILKVIPRDRKTFLFSATMTKKVQKLQRAALKNPVKCAVSSKYQTVEKLQQ
PRD   hhcchhhhhhhhhhhccccchhhhhhhhhccchhhhhhhhhhhccceeeeeeecccccchhhhh

SEQ   YYIFIPSKFKDTYLVYILNELAGNSFMIFCSTCNNTQRTALLLRNLGFTAIPLHGQMSQS
PRD   hhhhhhhhhhhhhhhhhhhhhhcccceeeeeeeecchhhhhhhhhhhhccccceeeccccchhh

SEQ   KRLGSLNKFKAKARSILLATDVASRGLDIPHVDVVVNFDIPTHSKDYIHRVGRTARAGRS
PRD   hhhhhhhhhhhhhhhcchhhhhhhhhcccccceeeeeeeccccccccceeeeeeccccccccc

SEQ   GKAITFVTQYDVELFQRIEHLIGKKLPGFPTQDDEVMMLTERVAEAQRFARMELREHGEK
PRD   cceeeeeeecchhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   KKRSREDAGDNDDTEGAIGVRNKVAGGKMKKRKGR
PRD   hhhhcccccccccccccccccccccccccccccc
```

Prosite for DKFZphfbr2_6o17.3

```
PS00001    274->278    ASN_GLYCOSYLATION    PDOC00001
PS00004    421->425    CAMP_PHOSPHO_SITE    PDOC00004
PS00005     25->28     PKC_PHOSPHO_SITE     PDOC00005
PS00005     72->75     PKC_PHOSPHO_SITE     PDOC00005
PS00005    209->212    PKC_PHOSPHO_SITE     PDOC00005
PS00005    229->232    PKC_PHOSPHO_SITE     PDOC00005
PS00005    276->279    PKC_PHOSPHO_SITE     PDOC00005
PS00005    300->303    PKC_PHOSPHO_SITE     PDOC00005
PS00005    354->357    PKC_PHOSPHO_SITE     PDOC00005
PS00005    360->363    PKC_PHOSPHO_SITE     PDOC00005
PS00005    400->403    PKC_PHOSPHO_SITE     PDOC00005
PS00006      9->13     CK2_PHOSPHO_SITE     PDOC00006
PS00006     25->29     CK2_PHOSPHO_SITE     PDOC00006
PS00006    186->190    CK2_PHOSPHO_SITE     PDOC00006
PS00006    368->372    CK2_PHOSPHO_SITE     PDOC00006
PS00006    391->395    CK2_PHOSPHO_SITE     PDOC00006
PS00006    424->428    CK2_PHOSPHO_SITE     PDOC00006
PS00008     66->72     MYRISTYL             PDOC00008
PS00008     71->77     MYRISTYL             PDOC00008
PS00008    116->122    MYRISTYL             PDOC00008
PS00008    120->126    MYRISTYL             PDOC00008
PS00008    128->134    MYRISTYL             PDOC00008
PS00009    382->386    AMIDATION            PDOC00009
PS00017     68->76     ATP_GTP_A            PDOC00017
PS00039    172->181    DEAD_ATP_HELICASE    PDOC00039
```

Pfam for DKFZphfbr2_6o17.3

```
HMM_NAME     DEAD and DEAH box helicases

HMM              *gLpPWILRnIyeMGFEkPTPIQQqAIPiILeGRDVMACAQTGSGKTAAF
                  G ++ ++++++++G++KPT+IQ +AIP++L+GRD+++ A TGSGKT+AF
Query         30  GVTDVLCEACDQLGWTKPTKIQIEAIPLALQGRDIIGLAETGSGKTGAF    78

HMM              lIPMLQHIDwdPWpqpPQdPrALILAPTRELAMQIQEEcRkFgkHMngIR
                 ++P+L  ++++P      + ++AL+L+PTRELA QI+E+++++G++++ ++
Query         79 ALPILNALLETP----QR-LFALVLTPTRELAFQISEQFEALGSSIG-VQ   122

HMM              ImcIYGGtnMRdQMRmLeRGpPHIVIATPGRLIDHIER.gtldLDrIeML
                 +++I+GG + + Q   L+++P HI+IATPGRLIDH+E+   ++L+++++L
Query        123 SAVIVGGIDSMSQSLALAKKP-HIIIATPGRLIDHLENTKGFNLRALKYL   171

HMM              VMDEADRMLDMGFIDQIRrIMrqIPMpwNRQTMMFSATMPdeIqELARrF
                 VMDEADR+L+M+F+ ++++I++ IP  ++R T +FSATM++++Q+L+R+
Query        172 VMDEADRILNMDFETEVDKILKVIP--RDRKTFLFSATMTKKVQKLQRAA   219

HMM              MRNPIRInIdMdElTtnEnIkQwYiyVerEMWKfdcLcrLIe*
                 ++NP+  ++ ++++T++  ++Q+YI+++  + K  +L+++++
Query        220 LKNPVKCAVSSKYQTVE-KLQQYYIFIP-SKFKDTYLVYILN    259


HMM_NAME     Helicases conserved C-terminal domain

HMM              *EileeWLknlGIrvmYIHGdMpQeERdeIMddFNnGEynVLIcTDVggR
```

```
                  ++ +  L+NLG++++ +HG+M+Q +R+  +++F++    +L++TDV++R
Query       277  QRTALLLRNLGFTAIPLHGQMSQSKRLGSLNKFKAKARSILLATDVASR   325

HMM             GIDIPdVNHVINYDMPWNPEqYIQRIGRTgRIG*
                G+DIP V++V+N+D+P ++ +YI+R+GRT+R+G
Query       326 GLDIPHVDVVVNFDIPTHSKDYIHRVGRTARAG    358
```

DKFZphfbr2_71o20
----

group: brain derived

DKFZphfbr2_71o20 encodes a novel 232 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits
on genomic level encoded by AC006186 (3 exons)

Sequenced by GBF

Locus: /map="10q22.1"

Insert length: 1768 bp
Poly A stretch at pos. 1742, polyadenylation signal at pos. 1726

```
   1 GGGGGCAGCA GGCCAAGGGG GAGGTGCGAG CGTGGACCTG GGACGGGTCT
  51 GGGCGGCTCT CGGTGGTTGG CACGGGTTCG CACACCCATT CAAGCGGCAG
 101 GACGCACTTG TCTTAGCAGT TCTCGCTGAC CGCGCTAGCT GCGGCTTCTA
 151 CGCTCCGGCA CTCTGAGTTC ATCAGCAAAC GCCCTGGCGT CTGTCCTCAC
 201 CATGCCTAGC CTTTGGGACC GCTTCTCGTC GTCGTCCACC TCCTCTTCGC
 251 CCTCGTCCTT GCCCCGAACT CCCACCCCAG ATCGGCCGCC GCGCTCAGCC
 301 TGGGGGTCGG CGACCCGGGA GGAGGGGTTT GACCGCTCCA CGAGCCTGGA
 351 GAGCTCGGAC TGCGAGTCCC TGGACAGCAG CAACAGTGGC TTCGGGCCGG
 401 AGGAAGACAC GGCTTACCTG GATGGGGTGT CGTTGCCCGA CTTCGAGCTG
 451 CTCAGTGACC CTGAGGATGA ACACTTGTGT GCCAACCTGA TGCAGCTGCT
 501 GCAGGAGAGC CTGGCCCAGG CGCGGCTGGG CTCTCGACGC CCTGCGCGCC
 551 TGCTGATGCC TAGCCAGTTG GTAAGCCAGG TGGGCAAAGA ACTACTGCGC
 601 CTGGCCTACA GCGAGCCGTG CGGCCTGCGG GGGGCGCTGC TGGACGTCTG
 651 CGTGGAGCAG GGCAAGAGCT GCCACAGCGT GGGCCAGCTG GCACTCGACC
 701 CCAGCCTGGT GCCCACCTTC CAGCTGACCC TCGTGCTGCG CCTGGACTCA
 751 CGACTCTGGC CCAAGATCCA GGGGCTGTTT AGCTCCGCCA ACTCTCCCTT
 801 CCTCCCTGGC TTCAGCCAGT CCCTGACGCT GAGCACTGGC TTCCGAGTCA
 851 TCAAGAAGAA GCTGTACAGC TCGGAACAGC TGCCCATTGA GGAGTGTTGA
 901 ACTTCAACCT GAGGGGGCCG ACAGTGCCCT CCAAGACAGA GACGACTGAA
 951 CTTTTGGGGT GGAGACTAGA GGCAGGAGCT GAGGGACTGA TTCCAGTGGT
1001 TGGAAAACTG AGGCAGCCAC CTAAAGTGGA GGTGGGGGAA TAGTGTTTCC
1051 CAGGAAGCTC ATTGAGTTGT GTGCGGGTGG CTCTGCATTG GGGACACATA
1101 CCCCTCAGTA CTGTAGCATG AAACAAAGGC TTAGGGGCCA ACAAGGCTTC
1151 CAGCTGGATG TGTGTGTAGC ATGTACCTTA TTATTTTTGT TACTGACAGT
1201 TAACAGTGGT GTGACATCCA GAGAGCAGCT GGGCTGCTCC CGCCCCCAGCC
1251 TGGCCCAGGG TGAAGGAAGA GGCACGTGCT CCTCAGAGCA GCCGGAGGGA
1301 AGGGGGAGGT CGGAGGTCGT GGAGGTGGTT TGTGTATCTT ACTGGTCTGA
1351 AGGGACCAAG TGTGTTTGTT GTTTGTTTTG TATCTTGTTT TTCTGATCGG
1401 AGCATCACTA CTGACCTGTT GTAGGCAGCT ATCTTACAGA CGCATGAATG
1451 TAAGAGTAGG AAGGGGTGGG TGTCAGGGAT CACTTGGGAT CTTTGACACT
1501 TGAAAAATTA CACCTGGCAG CTGCGTTTAA GCCTTCCCCC ATCGTGTACT
1551 GCAGAGTTGA GCTGGCAGGG GAGGGGCTGA GAGGGTGGGG GCTGGAACCC
1601 CTTCCCGGGA GGAGTGCCAT CTGGGTCTTC CATCTAGAAC TGTTTACATG
1651 AAGATAAGAT ACTCACTGTT CATGAATACA CTTGATGTTC AAGTATTAAG
1701 ACCTATGCAA TATTTTTTAC TTTTCTAATA AACATGTTTG TTAAAACAAA
1751 AAAAAAAAAA AAAAAAA
```

BLAST Results

Entry AC006186 from database EMBLNEW:
*** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 10 clone CRI-JC2048 map 10q22.1; HTGS phase 1, 4 unordered pieces.
Score = 6512, P = 0.0e+00, identities = 1326/1345
3 exons

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 202 bp to 897 bp; peptide length: 232
Category: putative protein

```
  1 MPSLWDRFSS SSTSSSPSSL PRTPTPDRPP RSAWGSATRE EGFDRSTSLE
 51 SSDCESLDSS NSGFGPEEDT AYLDGVSLPD FELLSDPEDE HLCANLMQLL
101 QESLAQARLG SRRPARLLMP SQLVSQVGKE LLRLAYSEPC GLRGALLDVC
151 VEQGKSCHSV GQLALDPSLV PTFQLTLVLR LDSRLWPKIQ GLFSSANSPF
201 LPGFSQSLTL STGFRVIKKK LYSSEQLPIE EC
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_71o20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_71o20, frame 1

Report for DKFZphfbr2_71o20.1

[LENGTH]	232	
[MW]	25354.60	
[pI]	4.87	
[PROSITE]	MYRISTYL	2
[PROSITE]	CK2_PHOSPHO_SITE	6
[PROSITE]	GLYCOSAMINOGLYCAN	1
[PROSITE]	PKC_PHOSPHO_SITE	1
[KW]	All_Alpha	
[KW]	LOW_COMPLEXITY	17.67 %

```
SEQ  MPSLWDRFSSSSTSSSPSSLPRTPTPDRPPRSAWGSATREEGFDRSTSLESSDCESLDSS
SEG  ........xxxxxxxxxxxxxxxxxxxxxx............xxxxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  NSGFGPEEDTAYLDGVSLPDFELLSDPEDEHLCANLMQLLQESLAQARLGSRRPARLLMP
SEG  xx..........................................................
PRD  cccccccccccccccccccceeeccccccchhhhhhhhhhhhhhhhhcccccccceeecc

SEQ  SQLVSQVGKELLRLAYSEPCGLRGALLDVCVEQGKSCHSVGQLALDPSLVPTFQLTLVLR
SEG  ............................................................
PRD  cccchhhhhhhhhhhcccccchhhhhhhhccccccccccccccccccccchhhhhhccc

SEQ  LDSRLWPKIQGLFSSANSPFLPGFSQSLTLSTGFRVIKKKLYSSEQLPIEEC
SEG  ....................................................
PRD  cccccccccccccccccccccccccceeeecccccccccccccccccccccc
```

Prosite for DKFZphfbr2_71o20.1

PS00002	62->66	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	111->114	PKC_PHOSPHO_SITE	PDOC00005
PS00006	3->7	CK2_PHOSPHO_SITE	PDOC00006
PS00006	38->42	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	85->89	CK2_PHOSPHO_SITE	PDOC00006
PS00008	141->147	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_71o20.1)

DKFZphfbr2_72b18
----

group: nucleic acid management

DKFZphfbr2_72b18 encodes a novel 715 amino acid protein with similarity to E. coli DNA-damage-inducibile protein dinP and other proteins induced by DNA-damage.

The novel protein is similar to dinP of E. coli, yqjH of B. subtilis, dinP of M. tuberculosis and T19K24.15 of A. thaliana. The dinB/P pathway is a second SOS-pathway in E. coli. Therefore the new gene seems to be involved in DNA repair.

The new protein can find application in modulating DNA repair and mutagenesis.

similarity to DNA damage induced genes

complete cDNA, complete cds, potential start at Bp 49, EST hits
localisation primer site B is missing!

Sequenced by LMU

Locus: /map="416.0 cR from top of Chr18 linkage group"??

Insert length: 2475 bp
Poly A stretch at pos. 2452, polyadenylation signal at pos. 2431

```
   1 GGGGGAGGAA GGCGGCGGCG ACGACGAGGA AGACGCCGAG GCCTGGGCCA
  51 TGGAACTGGC GGACGTGGGG GCGGCAGCCA GCTCGCAGGG AGTTCATGAT
 101 CAAGTGTTGC CCACACCAAA TGCTTCATCC AGAGTCATAG TACATGTGGA
 151 TCTGGATTGC TTTTATGCAC AAGTAGAAAT GATCTCAAAT CCAGAGCTAA
 201 AAGACAAACC TTTAGGGGTT CAACAGAAAT ATTTGGTGGT TACCTGCAAC
 251 TATGAAGCTA GGAAACTTGG AGTTAAGAAA CTTATGAATG TCAGAGATGC
 301 AAAAGAAAAG TGTCCACAGT TGGTATTAGT TAATGGAGAA GACCTGACCC
 351 GCTACAGAGA AATGTCTTAT AAGGTTACAG AATTACTGGA AGAATTTAGT
 401 CCAGTTGTTG AGAGACTTGG ATTTGATGAA AATTTTGTGG ATCTAACAGA
 451 AATGGTTGAG AAGAGACTAC AGCAGCTGCA AAGTGATGAA CTTTCTGCGG
 501 TGACTGTGTC GGGTCATGTA TACAATAATC AGTCTATAAA CCTGCTTGAC
 551 GTCTTGCACA TCAGACTACT TGTTGGATCT CAGATTGCAG CAGAGATGCG
 601 GGAAGCCATG TATAATCAGT TGGGGCTCAC TGGCTGTGCT GGAGTGGCTT
 651 CTAATAAACT GTTGGCAAAA TTAGTTTCTG GTGTCTTTAA ACCAAATCAA
 701 CAAACAGTCT TATTACCTGA AAGTTGTCAA CATCTTATTC ATAGTTTGAA
 751 TCACATAAAG GAAATACCTG GTATTGGCTA TAAAACTGCC AAATGTCTTG
 801 AAGCACTGGG TATCAATAGT GTGCGTGATC TCCAAACCTT TTCACCCAAA
 851 ATTTTAGAAA AAGAATTAGG AATTTCAGTT GCTCAGCGTA TCCAAAAGCT
 901 CAGTTTTGGA GAGGATAACT CCCCTGTGAT ACTCTCAGGA CCACCTCAGT
 951 CCTTTAGTGA AGAAGATTCA TTTAAAAAAT GTACATCTGA AGTTGAAGCT
1001 AAAAATAAGA TTGAAGAACT ACTTGCTAGT CTTTTAAACA GAGTATGCCA
1051 AGATGGAAGG AAGCCTCATA CAGTGAGATT AATAATCCGT CGGTATTCCT
1101 CTGAGAAGCA CTATGGTCGT GAGAGTCGTC AGTGCCCTAT TCCTTCACAT
1151 GTAATTCAGA AATTAGGGAC AGGAAATTAT GATGTGATGA CCCCAATGGT
1201 TGATATACTT ATGAAACTTT TTCGAAATAT GGTGAATGTG AAGATGCCAT
1251 TTCACCTTAC CCTTCTAAGT GTGTGCTTCT GCAACCTTAA AGCACTAAAT
1301 ACTGCTAAGA AAGGGCTTAT TGATTATTAT TTAATGCCAT CATTATCAAC
1351 TACTTCACGC TCTGGCAAGC ACAGTTTTAA AATGAAAGAC ACTCATATGG
1401 AAGATTTTCC CAAAGACAAA GAAACAAACC GGGATTTCCT ACCAAGTGGA
1451 AGAATTGAAA GTACAAGAAC TAGGGAGTCT CCACTAGATA CCACAAATTT
1501 TTCTAAAGAA AAAGACATTA ATGAATTCCC ACTCTGTTCA CTTCCTGAAG
1551 GTGTTGACCA AGAAGTCTCC AAGCAGCTTC CAGTAGATAT TCAAGAAGAA
1601 ATCCTTTCTG GAAAATCTAG GGAAAAAATTT CAAGGGAAAG GAAGTGTGAG
1651 TTGTCCATTA CATGCCTCTA GAGGAGTATT ATCTTTCTTT TCTAAAAAAC
1701 AAATGCAAGA TATTCCCATA AATCCTAGAG ATCATTTATC CAGTAGCAAA
1751 CAGGTATCCT CTGTATCTCC TTGTGAACCG GGAACATCAG GCTTTAATAG
1801 CAGTAGTTCT TCTTACATGT CTAGCCAAAA GGATTATTCA TATTATTTAG
1851 ATAATAGATT AAAAGATGAA CGAATAAGTC AAGGACCTAA AGAACCTCAA
1901 GGATTCCACT TTACAAATTC AAACCCTGCT GTGTCTGCTT TTCATTCATT
1951 TCCAAACTTG CAGAGTGAGC AACTTTTCTC CAGAAACCAC ACTACAGATA
2001 GCCATAAGCA AACAGTAGCA ACAGACTCTC ATGAAGGACT TACAGAAAAT
2051 AGAGAGCCAG ATTCTGTTGA TGAGAAAATT ACTTTCCCTT CTGACATTGA
2101 TCCTCAAGTT TTCTATGAAC TACCAGAAGC AGTACAAAAG GAACTGCTGG
2151 CAGAGTGGAA GAGAACAGGA TCAGATTTCC ACATTGGACA TAAATAAGCA
2201 TATTCAGCAA AAAGGTCTGA AAAGCAAGGG AATACCATTA TTTTCGGATT
2251 AGCGGTTTAT TAAGCTCTTC TATATTAAAC ACTAATAGAT ATTCAATAAC
2301 GGAGTAAACT GTTCCAGATA AAGCAAGAAT AGTTGCAAGA AGTAAATTCT
2351 GGCACAAAGC GTAAAAATAT AACAGAAGAA ATAATGTAAA ATACTATCTT
2401 TTATGTCTAA AGCCATTTTA TATTACTTTT CAATAAAAAG AATATCATGG
2451 TCAAAAAAAA AAAAAAAAAA AAAAC
```

BLAST Results

Entry HS086339 from database EMBL:
human STS WI-11064.
Score = 1523, P = 3.0e-64, identities = 327/343

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 50 bp to 2194 bp; peptide length: 715
Category: similarity to known protein

```
  1 MELADVGAAA SSQGVHDQVL PTPNASSRVI VHVDLDCFYA QVEMISNPEL
 51 KDKPLGVQQK YLVVTCNYEA RKLGVKKLMN VRDAKEKCPQ LVLVNGEDLT
101 RYREMSYKVT ELLEEFSPVV ERLGFDENFV DLTEMVEKRL QQLQSDELSA
151 VTVSGHVYNN QSINLLDVLH IRLLVGSQIA AEMREAMYNQ LGLTGCAGVA
201 SNKLLAKLVS GVFKPNQQTV LLPESCQHLI HSLNHIKEIP GIGYKTAKCL
251 EALGINSVRD LQTFSPKILE KELGISVAQR IQKLSFGEDN SPVILSGPPQ
301 SFSEEDSFKK CTSEVEAKNK IEELLASLLN RVCQDGRKPH TVRLIIRRYS
351 SEKHYGRESR QCPIPSHVIQ KLGTGNYDVM TPMVDILMKL FRNMVNVKMP
401 FHLTLLSVCF CNLKALNTAK KGLIDYYLMP SLSTTSRSGK HSFKMKDTHM
451 EDFPKDKETN RDFLPSGRIE STRTRESPLD TTNFSKEKDI NEFPLCSLPE
501 GVDQEVSKQL PVDIQEEILS GKSREKFQGK GSVSCPLHAS RGVLSFFSKK
551 QMQDIPINPR DHLSSSKQVS SVSPCEPGTS GFNSSSSSYM SSQKDYSYYL
601 DNRLKDERIS QGPKEPQGFH FTNSNPAVSA FHSFPNLQSE QLFSRNHTTD
651 SHKQTVATDS HEGLTENREP DSVDEKITFP SDIDPQVFYE LPEAVQKELL
701 AEWKRTGSDF HIGHK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72b18, frame 2

PIR:H64747 DNA-damage-inducibile protein dinP - Escherichia coli, N = 2, Score = 212, P = 4.2e-27

PIR:H69963 DNA-damage repair protein homolog yqjH - Bacillus subtilis, N = 2, Score = 230, P = 5.2e-26

>PIR:H69963 DNA-damage repair protein homolog yqjH - Bacillus subtilis
Length = 414

HSPs:

Score = 230 (34.5 bits), Expect = 5.2e-26, Sum P(2) = 5.2e-26
Identities = 47/112 (41%), Positives = 73/112 (65%)

```
Query:    27 SRVIVHVDLDCFYAQVEMISNPELKDKPLGV-----QQKYLVVTCNYEARKLGVKKLMNV 81
             SR+I H+D++ FYA VEM  +P L+ KP+ V      ++K +VVTC+YEAR  GVK  M V
Sbjct:     5 SRIIFHIDMNSFYASVEMAYDPALRGKPVAVAGNVKERKGIVVTCSYEARARGVKTTMPV 64

Query:    82 RDAKEKCPQLVLVNGEDLTRYREMSYKVTELLEEFSPVVERLGFDENFVDLTE 134
               AK  CP+L+++    +  RYR  S  +  +L E++ +VE +  DE ++D+T+
Sbjct:    65 WQAKRHCPELIVLP-PNFDRYRNSSRAMFTILREYTDLVEPVSIDEGYMDMTD 116
```

Score = 137 (20.6 bits), Expect = 5.2e-26, Sum P(2) = 5.2e-26
Identities = 43/148 (29%), Positives = 75/148 (50%)

```
Query:   178 QIAAEMREAMYNQLGLTGCAGVASNKLLAKLVSGVFKPNQQTVLLPESCQHLIHSLNHIK 237
             + A E++  +  +L L     G+A NK LAK+ S + KP    T+L         ++  L  +
Sbjct:   125 ETAKEIQSRLQKELLLPSSIGIAPNKFLAKMASDMKKPLGITILRKRQVPDILWPLP-VG 183

Query:   238 EIPGIGYKTAKCLEALGINSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSG 297
             E+ G+G KTA+ L+ LGI+++ +L          L++ LGI+  R++  + G  ++PV
Sbjct:   184 EMHGVGKKTAEKLKGLGIHTIGELAAADEHSLKRLLGIN-GPRLKNKANGIHHAPV---- 238

Query:   298 PPQSFSEEDSFKKCTSEVEAKNKIEELL 325
              P+   E  S    ++       +  EELL
```

Sbjct: 239 DPERIYEFKSVGNSSTLSHDSSDEEELL 266

Pedant information for DKFZphfbr2_72b18, frame 2

Report for DKFZphfbr2_72b18.2

```
[LENGTH]        715
[MW]            80300.63
[pI]            6.37
[HOMOL]         TREMBL:SPBC16A3_11 gene: "SPBC16A3.11"; product: "hypothetical protein";
S.pombe chromosome II cosmid c16A3. 5e-30
[FUNCAT]        11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair)         [S. cerevisiae, YDR419w] 2e-15
[FUNCAT]        1 genome replication, transcription, recombination and repair      [M.
genitalium, MG360] 3e-13
[PIRKW]         SOS mutagenesis 2e-11
[PIRKW]         DNA repair 2e-11
[PIRKW]         induced mutagenesis 2e-11
[SUPFAM]        umuC protein 3e-29
[PROSITE]       MYRISTYL        6
[PROSITE]       AMIDATION       1
[PROSITE]       CAMP_PHOSPHO_SITE       2
[PROSITE]       CK2_PHOSPHO_SITE        15
[PROSITE]       PROKAR_LIPOPROTEIN      1
[PROSITE]       TYR_PHOSPHO_SITE        2
[PROSITE]       PKC_PHOSPHO_SITE        21
[PROSITE]       ASN_GLYCOSYLATION       5
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY      4.20 %

SEQ     MELADVGAAASSQGVHDQVLPTPNASSRVIVHVDLDCFYAQVEMISNPELKDKPLGVQQK
SEG     ............................................................
PRD     ccceeeeeeecccccceeeccccccceeeeeeeccchhhhhhhhhccccccccceeeecc

SEQ     YLVVTCNYEARKLGVKKLMNVRDAKEKCPQLVLVNGEDLTRYREMSYKVTELLEEFSPVV
SEG     ............................................................
PRD     ceeeehhhhhhhhhhccchhhhhhhhcccceeeeccccccchhhhhhhhhhhhhhhccce

SEQ     ERLGFDENFVDLTEMVEKRLQQLQSDELSAVTVSGHVYNNQSINLLDVLHIRLLVGSQIA
SEG     ............................................................
PRD     eeecccchhhhhhhhhhhhhhhhhhcccccceeeeeccccccchhhhhhhhhhhhhhhhh

SEQ     AEMREAMYNQLGLTGCAGVASNKLLAKLVSGVFKPNQQTVLLPESCQHLIHSLNHIKEIP
SEG     ............................................................
PRD     hhhhhhhhhhcceeeecccchhhhhhhhhhhhhccccceeeeecchhhhhhhhhcccccc

SEQ     GIGYKTAKCLEALGINSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSGPPQ
SEG     ............................................................
PRD     ccchhhhhhhhhhcccccchhhhhhhhhhhhhhcchhhhhhhhhhcccccceeeeccccc

SEQ     SFSEEDSFKKCTSEVEAKNKIEELLASLLNRVCQDGRKPHTVRLIIRRYSSEKHYGRESR
SEG     ............................................................
PRD     ccccccccccchhhhhhhhhhhhhhhhhhhhhhhcccccccceeeehhhhhhhhhhhcccc

SEQ     QCPIPSHVIQKLGTGNYDVMTPMVDILMKLFRNMVNVKMPFHLTLLSVCFCNLKALNTAK
SEG     ............................................................
PRD     cccccceeeeccccccccchhhhhhhhhhhhhhhhhhccccceeeeeeeeechhhhhhhh

SEQ     KGLIDYYLMPSLSTTSRSGKHSFKMKDTHMEDFPKDKETNRDFLPSGRIESTRTRESPLD
SEG     ............................................................
PRD     hhhheeeecccccccccccccceeeccccccccccccccccccccccccccccccccccc

SEQ     TTNFSKEKDINEFPLCSLPEGVDQEVSKQLPVDIQEEILSGKSREKFQGKGSVSCPLHAS
SEG     ............................................................
PRD     cccccccccccccccccchhhhhhhhhhhhhhhhhhhhccccceeeeecccccccchhhh

SEQ     RGVLSFFSKKQMQDIPINPRDHLSSSKQVSSVSPCEPGTSGFNSSSSSYMSSQKDYSYYL
SEG     .......................xxxxxxxxxx......xxxxxxxxxxxxxxxxxxxx.
PRD     hccccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhh

SEQ     DNRLKDERISQGPKEPQGFHFTNSNPAVSAFHSFPNLQSEQLFSRNHTTDSHKQTVATDS
SEG     ............................................................
PRD     hhhhhhhhhhcccccccceeeecccccceeecccccccchhhhhhhcccccccceeeeeecc

SEQ     HEGLTENREPDSVDEKITFPSDIDPQVFYELPEAVQKELLAEWKRTGSDFHIGHK
SEG     .......................................................
PRD     ccccccccccccccccccccccccceeehhhhhhhhhhhhhhhhhcccccccccc
```

Prosite for DKFZphfbr2_72b18.2

PS00001	24->28	ASN_GLYCOSYLATION	PDOC00001
PS00001	160->164	ASN_GLYCOSYLATION	PDOC00001
PS00001	483->487	ASN_GLYCOSYLATION	PDOC00001
PS00001	583->587	ASN_GLYCOSYLATION	PDOC00001
PS00001	646->650	ASN_GLYCOSYLATION	PDOC00001
PS00004	309->313	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	347->351	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	106->109	PKC_PHOSPHO_SITE	PDOC00005
PS00005	201->204	PKC_PHOSPHO_SITE	PDOC00005
PS00005	246->249	PKC_PHOSPHO_SITE	PDOC00005
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	265->268	PKC_PHOSPHO_SITE	PDOC00005
PS00005	307->310	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	418->421	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	438->441	PKC_PHOSPHO_SITE	PDOC00005
PS00005	442->445	PKC_PHOSPHO_SITE	PDOC00005
PS00005	459->462	PKC_PHOSPHO_SITE	PDOC00005
PS00005	466->469	PKC_PHOSPHO_SITE	PDOC00005
PS00005	471->474	PKC_PHOSPHO_SITE	PDOC00005
PS00005	520->523	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	565->568	PKC_PHOSPHO_SITE	PDOC00005
PS00005	592->595	PKC_PHOSPHO_SITE	PDOC00005
PS00005	651->654	PKC_PHOSPHO_SITE	PDOC00005
PS00006	46->50	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	285->289	CK2_PHOSPHO_SITE	PDOC00006
PS00006	301->305	CK2_PHOSPHO_SITE	PDOC00006
PS00006	303->307	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	448->452	CK2_PHOSPHO_SITE	PDOC00006
PS00006	459->463	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	497->501	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	592->596	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	681->685	CK2_PHOSPHO_SITE	PDOC00006
PS00006	706->710	CK2_PHOSPHO_SITE	PDOC00006
PS00007	101->108	TYR_PHOSPHO_SITE	PDOC00007
PS00007	348->356	TYR_PHOSPHO_SITE	PDOC00007
PS00008	7->13	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00008	192->198	MYRISTYL	PDOC00008
PS00008	198->204	MYRISTYL	PDOC00008
PS00008	274->280	MYRISTYL	PDOC00008
PS00008	663->669	MYRISTYL	PDOC00008
PS00009	335->339	AMIDATION	PDOC00009
PS00013	186->197	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphfbr2_72b18.2)

DKFZphfbr2_72d13

group: brain derived

DKFZphfbr2_72d13 encodes a novel 165 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

seems to be testis specific 9 of 10 EST hits are from testis librarys

Sequenced by LMU

Locus: unknown

Insert length: 723 bp
Poly A stretch at pos. 704, no polyadenylation signal found

```
  1 AGGGGGGGTA TGGGGGAGGG GGAGACTCTG CAGGAGCCTA ATTCCCCACT
 51 CTGAGCTCAC CCTTCTGTCT GCCCGGGCCC TACCCCTTCC CCTACTCTCA
101 CCCTTATAAT CCTTTTCAGC ACTAGGTCTT CCCGTCACCT CCACCTCTCT
151 CCATGACCCG GCTCTGCTTA CCCAGACCCG AAGCACGTGA GGATCCGATC
201 CCAGTTCCTC CAAGGGGCCT GGGTGCTGGG GAGGGGTCAG GTAGTCCAGT
251 GCGTCCACCT GTATCCACCT GGGGCCCTAG CTGGGCCCAG CTCCTGGACA
301 GTGTCCTATG GCTGGGGGCA CTAGGACTGA CAATCCAGGC AGTCTTTTCC
351 ACCACTGGCC CAGCCCTGCT GCTGCTTCTG GTCAGCTTCC TCACCTTTGA
401 CCTGCTCCAT AGGCCCGCAG GTCACACTCT GCCACAGCGC AAACTTCTCA
451 CCAGGGGCCA GAGTCAGGGG GCCGGTGAAG GTCCTGGACA GCAGGAGGCT
501 CTACTCCTGC AAATGGGTAC AGTCTCAGGA CAACTTAGCC TCCAGGACGC
551 ACTGCTGCTG CTGCTCATGG GGCTGGGCCC GCTCCTGAGA GCCTGTGGCA
601 TGCCCTTGAC CCTGCTTGGC CTGGCTTTCT GCCTCCATCC TTGGGCCTGA
651 GAGCCCCTCC CCACAACTCA GTGTCCTTCA AATATACAAT GACCACCCTT
701 CTTCAAAAAA AAAAAAAAAA AAC
```

BLAST Results

Entry HS860F19 from database EMBLNEW:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 860F19
Score = 2059, P = 1.1e-85, identities = 423/434
2 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 153 bp to 647 bp; peptide length: 165
Category: putative protein
Classification: no clue

```
  1 MTRLCLPRPE AREDPIPVPP RGLGAGEGSG SPVRPPVSTW GPSWAQLLDS
 51 VLWLGALGLT IQAVFSTTGP ALLLLLVSFL TFDLLHRPAG HTLPQRKLLT
101 RGQSQGAGEG PGQQEALLLQ MGTVSGQLSL QDALLLLLMG LGPLLRACGM
151 PLTLLGLAFC LHPWA
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72d13, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_72d13, frame 3

Report for DKFZphfbr2_72d13.3

```
[LENGTH]      165
[MW]          17393.73
[pI]          7.80
[BLOCKS]      BL00068A Malate dehydrogenase proteins
[KW]          TRANSMEMBRANE  2
[KW]          LOW_COMPLEXITY    29.70 %

SEQ   MTRLCLPRPEAREDPIPVPPRGLGAGEGSGSPVRPPVSTWGPSWAQLLDSVLWLGALGLT
SEG   ............................................................
PRD   cccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhcccccc
MEM   ............................................................

SEQ   IQAVFSTTGPALLLLLVSFLTFDLLHRPAGHTLPQRKLLTRGQSQGAGEGPGQQEALLLQ
SEG   ...........xxxxxxxxxxxxxx................xxxxxxxxxxxxxxx....
PRD   eeeecccccchhhhhhhhhhhhhhhcccccccccccccccccccccccccccchhhhhhh
MEM   ........MMMMMMMMMMMMMMMMM...................................

SEQ   MGTVSGQLSLQDALLLLLMGLGPLLRACGMPLTLLGLAFCLHPWA
SEG   .....xxxxxxxxxxxxxxxxxxxx....................
PRD   hcccccchhhhhhhhhhhhhcchhhhhcccccchhhhhhhccccc
MEM   .............MMMMMMMMMMMMMMMMMM..............
```

(No Prosite data available for DKFZphfbr2_72d13.3)

(No Pfam data available for DKFZphfbr2_72d13.3)

DKFZphfbr2_72112
----

group: nucleic acid management

Summary DKFZphfbr2_72112 encodes a novel 344 amino acid protein with similarity to YDR126w and other S. cerevisiae proteins.

The novel protein contains a myc-type, helix-loop-helix dimerization domain signature. This helix-loop-helix domain mediates protein dimerization and has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers. Therefore, the protein could be a novel DNA-binding protein.

The new protein can application in modulating gene expression.

similarity to YDR126w ;
membrane regions: 2

similarity to YDR126w

complete cDNA complete cds, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 1270 bp
Poly A stretch at pos. 1251, no polyadenylation signal found

```
   1 GGGGGCGCCC GGGAGGCGCC GGAGCCCAGC GGCTGGCGCC AGATCCAGGC
  51 TCCTGGAAGA ACCATGTCCG GCAGCTACTG GTCATGCCAG GCACACACTG
 101 CTGCCCAAGA GGAGCTGCTG TTTGAATTAT CTGTGAATGT TGGGAAGAGG
 151 AATGCCAGAG CTGCCGGCTG AAAATTACCC AACCAAGAGA AATCTGCAGG
 201 ATGGACTTTC TGGTCCTCTT CTTGTTCTAC CTGGCTTCGG TGCTGATGGG
 251 TCTTGTTCTT ATCTGCGTCT GCTCGAAAAC CCATAGCTTG AAAGGCCTGG
 301 CCAGGGGAGG AGCACAGATA TTTTCCTGTA TAATTCCAGA ATGTCTTCAG
 351 AGAGCCGTGC ATGGATTGCT TCATTACCTT TTCCATACGA GAAACCACAC
 401 CTTCATTGTC CTGCACCTGG TCTTGCAAGG GATGGTTTAT ACTGAGTACA
 451 CCTGGGAAGT ATTTGGCTAC TGTCAGGAGC TGGAGTTGTC CTTGCATTAC
 501 CTTCTTCTGC CCTATCTGCT GCTAGGTGTA AACCTGTTTT TTTTCACCCT
 551 GACTTGTGGA ACCAATCCTG GCATTATAAC AAAAGCAAAT GAATTATTAT
 601 TTCTTCATGT TTATGAATTT GATGAAGTGA TGTTTCCAAA GAACGTGAGG
 651 TGCTCTACTT GTGATTTAAG GAAACCAGCT CGATCCAAGC ACTGCAGTGT
 701 GTGTAACTGG TGTGTGCACC GTTTCGACCA TCACTGTGTT TGGGTGAACA
 751 ACTGCATCGG GGCCTGGAAC ATCAGGTACT TCCTCATCTA CGTCTTGACC
 801 TTGACGGCCT CGGCTGCCAC CGTCGCCATT GTGAGCACCA CTTTTCTGGT
 851 CCACTTGGTG GTGATGTCAG ATTTATACCA GGAGACTTAC ATCGATGACC
 901 TTGGACACCT CCATGTTATG GACACGGTCA TTCTTATTCA GTACCTGTTC
 951 CTGACTTTTC CACGGATTGT CTTCATGCTG GGCTTTGTCG TGGTCCTGAG
1001 CTTCCTCCTG GGTGGCTACC TGTTGTCTGT CCTGTATCTG GCGGCCACCA
1051 ACCAGACTAC TAACGAGTGG TACAGAGGTG TCTGGGCCTG GTGCCAGCGT
1101 TGTCCCCTTG TGGCCTGGCC TCCGTCAGCA GAGCCCCAAG TCCACCGGAA
1151 CATTCACTCC CATGGGCTTC GGAGCAACCT TCAAGAGATC TTTCTACCTG
1201 CCTTTCCATG TCATGAGAGG AAGAAACAAG AATGACAAGT GTATGACTGC
1251 CAAAAAAAAA AAAAAAAAAC
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 201 bp to 1232 bp; peptide length: 344
Category: similarity to unknown protein

```
  1 MDFLVLFLFY LASVLMGLVL ICVCSKTHSL KGLARGGAQI FSCIIPECLQ
 51 RAVHGLLHYL FHTRNHTFIV LHLVLQGMVY TEYTWEVFGY CQELELSLHY
101 LLLPYLLLGV NLFFFTLTCG TNPGIITKAN ELLFLHVYEF DEVMFPKNVR
151 CSTCDLRKPA RSKHCSVCNW CVHRFDHHCV WVNNCIGAWN IRYFLIYVLT
201 LTASAATVAI VSTTFLVHLV VMSDLYQETY IDDLGHLHVM DTVILIQYLF
251 LTFPRIVFML GFVVVLSFLL GGYLLSVLYL AATNQTTNEW YRGVWAWCQR
301 CPLVAWPPSA EPQVHRNIHS HGLRSNLQEI FLPAFPCHER KKQE
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72112, frame 3

TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein"; S.pombe chromosome II cosmid c13G1., N = 2, Score = 247, P = 1.4e-22

TREMBL:CED2021_3 gene: "D2021.2"; Caenorhabditis elegans cosmid D2021., N = 1, Score = 209, P = 9e-17

TREMBL:CEC43H6_2 gene: "C43H6.7"; Caenorhabditis elegans cosmid C43H6., N = 1, Score = 206, P = 5.2e-15

PIR:S52691 probable membrane protein YDR126w - yeast (Saccharomyces cerevisiae), N = 1, Score = 207, P = 8.4e-15

PIR:E71607 metal binding protein (DHHC domain) PFB0725c - malaria parasite (Plasmodium falciparum), N = 1, Score = 182, P = 1.1e-13

```
>TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein";
     S.pombe chromosome II cosmid c13G1.
           Length = 356

  HSPs:

 Score = 247 (37.1 bits), Expect = 1.4e-22, Sum P(2) = 1.4e-22
 Identities = 55/148 (37%), Positives = 85/148 (57%)

Query:    52 AVHGLLHYLFHTRNH--TFIVLHLVLQGM----VYTEYTWEVFGYCQELELSLHYLLLPY 105
             A+  L +Y+ +  N     F+ L L+  G+     +Y      + F      + + L  +LLPY
Sbjct:    64 AMRSLSNYVLYKNNPLVVFLYLALITIGIASFFIYGSSLTQKFSIIDWISV-LTSVLLPY 122

Query:   106 LLLGVNLFFFTLTCGTNPGIITKANELLFLHVYEFD-EVMFPKNVRCSTCDLRKPARSKH 164
                 ++L+    +    +NPG I   N       + +D ++ FP    +CSTC    KPARSKH
Sbjct:   123 ----ISLY---IAAKSNPGKIDLKNWNEASRRFPYDYKIFFPN--KCSTCKFEKPARSKH 173

Query:   165 CSVCNWCVHRFDHHCVWVNNCIGAWNIRYFLIYVL 199
             C +CN CV +FDHHC+W+NNC+G  N RYF +++L
Sbjct:   174 CRLCNICVEKFDHHCIWINNCVGLNNARYFFLFLL 208

 Score = 43 (6.5 bits), Expect = 1.4e-22, Sum P(2) = 1.4e-22
 Identities = 10/35 (28%), Positives = 17/35 (48%)

Query:   257 VFMLGFVV-VLSFLLGGYLLSVLYLAATNQTTNEW 290
             VF++   +  VL    L GY   ++Y   T    + +W
Sbjct:   254 VFLISLICSVLVLCLLGYEFFLVYAGYTTNESEKW 288
```

Pedant information for DKFZphfbr2_72112, frame 3

Report for DKFZphfbr2_72112.3

[LENGTH] 344
[MW] 39677.23
[pI] 7.26
[HOMOL] TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein"; S.pombe chromosome II cosmid c13G1. 3e-17
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YDR126w] 1e-16
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YDR264c] 8e-05
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YDR264c] 8e-05
[PIRKW] transmembrane protein 4e-15
[SUPFAM] ankyrin repeat homology 1e-10
[SUPFAM] unassigned ankyrin repeat proteins 1e-10
[PROSITE] MYRISTYL 4
[PROSITE] CK2_PHOSPHO_SITE 3

```
[PROSITE]     PKC_PHOSPHO_SITE       1
[PROSITE]     ASN_GLYCOSYLATION      2
[KW]          SIGNAL_PEPTIDE 30
[KW]          TRANSMEMBRANE  2
[KW]          LOW_COMPLEXITY     16.57 %
```

```
SEQ   MDFLVLFLFYLASVLMGLVLICVCSKTHSLKGLARGGAQIFSCIIPECLQRAVHGLLHYL
SEG   ............................................................
PRD   ccchhhhhhhhhhhhhheeeeeecccccеeeeeecccccеeeeeeeehhhhhhhhhheee
MEM   ............................................................

SEQ   FHTRNHTFIVLHLVLQGMVYTEYTWEVFGYCQELELSLHYLLLPYLLLGVNLFFFTLTCG
SEG   ................................xxxxxxxxxxxxxxxxxxx.........
PRD   ecccchhhhhhhhhhccchhhhhhhhheeeeccceeehhhhhhhhhhhhhhcccceeeeecc
MEM   .................................MMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ   TNPGIITKANELLFLHVYEFDEVMFPKNVRCSTCDLRKPARSKHCSVCNWCVHRFDHHCV
SEG   ............................................................
PRD   ccccccccccchhhhhhhhcccccccceeeecccccccccccccccccceeeecccccccc
MEM   M.............MMMMMMMMMMMMMMMMMMMMMMMMMMM...................

SEQ   WVNNCIGAWNIRYFLIYVLTLTASAATVAIVSTTFLVHLVVMSDLYQETYIDDLGHLHVM
SEG   .................xxxxxxxxxxxxxxxxx..........................
PRD   cccccccccccchhhhhhhhccchhhhhhhhhhhhhhhhhcccccccccccccccccchh
MEM   ............................................................

SEQ   DTVILIQYLFLTFPRIVFMLGFVVVLSFLLGGYLLSVLYLAATNQTTNEWYRGVWAWCQR
SEG   ...................xxxxxxxxxxxxxxxxxxxx.....................
PRD   hhhhhhhhhhhhhhhhccccccccceeecccchhhhhhhhhcccchhhhhhhhhhhcccc
MEM   ............................................................

SEQ   CPLVAWPPSAEPQVHRNIHSHGLRSNLQEIFLPAFPCHERKKQE
SEG   ............................................
PRD   cccccccccccccceeeccccccccccccceeeeeccccccccc
MEM   ............................................
```

Prosite for DKFZphfbr2_72112.3

PS00001	65->69	ASN_GLYCOSYLATION	PDOC00001
PS00001	284->288	ASN_GLYCOSYLATION	PDOC00001
PS00005	29->32	PKC_PHOSPHO_SITE	PDOC00005
PS00006	152->156	CK2_PHOSPHO_SITE	PDOC00006
PS00006	229->233	CK2_PHOSPHO_SITE	PDOC00006
PS00006	286->290	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	77->83	MYRISTYL	PDOC00008
PS00008	120->126	MYRISTYL	PDOC00008
PS00008	322->328	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_72112.3)

DKFZphfbr2_72m16

group: unknown

DKFZphfbr2_72m16 encodes a novel 287 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="26.2 cR from top of Chr16 linkage group"

Insert length: 1462 bp
Poly A stretch at pos. 1441, polyadenylation signal at pos. 1421

```
   1 GGGGAGGACC GGAGGACCGA GGACAGAAAG ATTGGTGGAC AGGAGCAGCG
  51 GCCGGTGGGG AGGGCGCTCG GCGGCGGCCT GCGGCCATGG CCACCGTGAT
 101 GGCAGCGACG GCGGCGGAGC GGGCGGTGCT GGAGGAGGAG TTCCGCTGGC
 151 TGCTGCACGA CGAGGTGCAC GCTGTGTTGA AGCAGCTGCA GGACATCCTC
 201 AAGGAGGCCT CTCTGCGCTT CACTCTGCCG GGCTCCGGCA CTGAGGGGCC
 251 CGCCAAGCAA GAGAACTTCA TCCTAGGCAG CTGTGGCACA GACCAGGTGA
 301 AGGGTGTGCT GACTCTGCAG GGGGATGCCC TCAGCCAGGC GGATGTGAAC
 351 CTGAAGATGC CCCGGAACAA CCAGCTGCTG CACTTCGCCT TCCGGGAGGA
 401 CAAGCAGTGG AAGCTGCAGC AGATCCAGGA TGCCAGAAAC CATGTGAGCC
 451 AAGCCATTTA CCTGCTTACC AGCCGGGACC AGAGCTACCA GTTCAAGACG
 501 GGCGCTGAGG TCCTCAAGCT GATGGACGCA GTGATGCTGC AGCTGACCAG
 551 AGCCCGAAAC CGGCTCACCA CCCCCGCCAC CCTCACCCTC CCCGAGATCG
 601 CCGCCAGCGG CCTCACGCGG ATGTTCGCCC CTGCCCTGCC GTCCGACCTG
 651 CTGGTCAACG TCTACATCAA CCTCAACAAG CTCTGCCTCA CGGTGTACCA
 701 GCTGCATGCC CTGCAGCCCA ACTCCACCAA GAACTTCCGC CCAGCTGGGG
 751 GCGCGGTGCT GCATAGCCCT GGGGCCATGT TCGAGTGGGG CTCTCAGCGC
 801 CTGGAGGTGA GCCACGTGCA CAAAGTGGAG TGCGTGATCC CCTGGCTCAA
 851 CGACGCCCTG GTCTACTTCA CCGTCTCCCT GCAGCTCTGC CAGCAGCTTA
 901 AGGACAAGAT CTCCGTGTTC TCCAGCTACT GGAGCTACAG ACCCTTCTGA
 951 TCACAGCACC CAGGAGCTTG TCTCCAGGAA GGCGGCCCCG TCCCCTACTC
1001 ATACCCACCA CAGAGCACCA GCCAGTGCCA ACGCCAGGCT GCTATTTATC
1051 TCCCTATCCC ACCCCCTACC CCACCTAACA CATTTGCACT GCCGGGAATG
1101 GACACTGGAA GTGCCAGGAG GAAGGAAGGC TGGTTTGGTG GGGTAGTGGG
1151 GAGGTCAGGG AGGCGGGGCC AAGGGTGTCC CACATTCCCA ACACCGCCCT
1201 CTGATCACCA TGGGAATCTT TGGACTCAGG ACAGGGCCAG GCGCAGGGCT
1251 CTCCCTCCTC TCCCCTTCGC TGTCCCCTCC CCCTGGAGGG CATGGTGTCG
1301 GGGGGTGGCA CTGAGCTATG AGTCCCGGGG ATGGTGAGGA ACGCCACAGA
1351 CAGAGCCACC CTAGGAGTGA GTATAGTGCT GGTGACTGTG TTTCATAGCC
1401 CCAGTCCAGG GCTGTCTAAG AAATAAAGAT CATCAGACTC CAAAAAAAAA
1451 AAAAAAAAAA AC
```

BLAST Results

Entry HS604351 from database EMBL:
human STS WI-18474.
Score = 1178, P = 1.5e-48, identities = 250/268

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 87 bp to 947 bp; peptide length: 287
Category: similarity to unknown protein

```
  1 MATVMAATAA ERAVLEEEFR WLLHDEVHAV LKQLQDILKE ASLRFTLPGS
 51 GTEGPAKQEN FILGSCGTDQ VKGVLTLQGD ALSQADVNLK MPRNNQLLHF
101 AFREDKQWKL QQIQDARNHV SQAIYLLTSR DQSYQFKTGA EVLKLMDAVM
151 LQLTRARNRL TTPATLTLPE IAASGLTRMF APALPSDLLV NVYINLNKLC
201 LTVYQLHALQ PNSTKNFRPA GGAVLHSPGA MFEWGSQRLE VSHVHKVECV
251 IPWLNDALVY FTVSLQLCQQ LKDKISVFSS YWSYRPF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72m16, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_72m16, frame 3

Report for DKFZphfbr2_72m16.3

```
[LENGTH]     287
[MW]         32254.40
[pI]         8.30
[HOMOL]      TREMBL:AF025459_2 gene: "H14A12.3"; Caenorhabditis elegans cosmid H14A12. 3e-14

[PROSITE]    MYRISTYL            1
[PROSITE]    CK2_PHOSPHO_SITE    6
[PROSITE]    PKC_PHOSPHO_SITE    5
[PROSITE]    ASN_GLYCOSYLATION   1
[KW]         Alpha_Beta
[KW]         LOW_COMPLEXITY      6.27 %

SEQ   MATVMAATAAERAVLEEEFRWLLHDEVHAVLKQLQDILKEASLRFTLPGSGTEGPAKQEN
SEG   xxxxxxxxxxxxxxxxxx..........................................
PRD   ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccchhhh

SEQ   FILGSCGTDQVKGVLTLQGDALSQADVNLKMPRNNQLLHFAFREDKQWKLQQIQDARNHV
SEG   ............................................................
PRD   hhccccccceeeeeeeecccccchhhhhhhcccccchhhhhhhhhchhhhhhhhhhhhchh

SEQ   SQAIYLLTSRDQSYQFKTGAEVLKLMDAVMLQLTRARNRLTTPATLTLPEIAASGLTRMF
SEG   ............................................................
PRD   hhhhhhhhcccccceeecchhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccc

SEQ   APALPSDLLVNVYINLNKLCLTVYQLHALQPNSTKNFRPAGGAVLHSPGAMFEWGSQRLE
SEG   ............................................................
PRD   cccccccceeeeehhhhhhhhhheeeeccccccccccccccceeeccccccccccccccee

SEQ   VSHVHKVECVIPWLNDALVYFTVSLQLCQQLKDKISVFSSYWSYRPF
SEG   ...............................................
PRD   eeeeeeeeeeeccccceeeeeeeehhhhhhhhhhhheeeeeeeeccc
```

Prosite for DKFZphfbr2_72m16.3

PS00001	212->216	ASN_GLYCOSYLATION	PDOC00001
PS00005	42->45	PKC_PHOSPHO_SITE	PDOC00005
PS00005	128->131	PKC_PHOSPHO_SITE	PDOC00005
PS00005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS00005	236->239	PKC_PHOSPHO_SITE	PDOC00005
PS00005	283->286	PKC_PHOSPHO_SITE	PDOC00005
PS00006	8->12	CK2_PHOSPHO_SITE	PDOC00006
PS00006	50->54	CK2_PHOSPHO_SITE	PDOC00006
PS00006	83->87	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	138->142	CK2_PHOSPHO_SITE	PDOC00006
PS00006	167->171	CK2_PHOSPHO_SITE	PDOC00006
PS00008	64->70	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_72m16.3)

DKFZphfbr2_72n12

group: brain derived

DKFZphfbr2_72n12 encodes a novel 117 amino acid protein with similarity to a protein with conserved sequence in bacteria and eukariota.

The novel protein is very similar to human MM46, human and rat ganglioside expression factor-2 (GEF2), C. elegans 14.8 kD protein C32D5.9 and Laccaria bicolor symbiosis-related protein LBU93506_1. The function of this highly conserved proteins is not known.

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to rat GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2)

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="12"

Insert length: 1880 bp
Poly A stretch at pos. 1859, polyadenylation signal at pos. 1830

```
   1 GGGGGCCGGT ATTTCTCCAT CTGGCTCTCC TCTACCTCCA GGCAGGCTCA
  51 CCCGAGATCC CCGCCCCGAA CCCCCCCTGC ACACTCGGCC CAGCGCTGTT
 101 GCCCCCGGAG CGGACGTTTC TGCAGCTATT CTGAGCACAC CTTGACGTCG
 151 GCTGAGGGAG CGGGACAGGG TCAGCGGCGA AGGAGGCAGG CCCCGCGCGG
 201 GGATCTCGGA AGCCCTGCGG TGCATCATGA AGTTCCAGTA CAAGGAGGAC
 251 CATCCCTTTG AGTATCGGAA AAAGGAAGGA GAAAAGATCC GGAAGAAATA
 301 TCCGGACAGG GTCCCCGTGA TTGTAGAGAA GGCTCCAAAA GCCAGGGTGC
 351 CTGATCTGGA CAAGAGGAAG TACCTAGTGC CCTCTGACCT TACTGTTGGC
 401 CAGTTCTACT TCTTAATCCG GAAGAGAATC CACCTGAGAC CTGAGGACGC
 451 CTTATTCTTC TTTGTCAACA ACACCATCCC TCCCACCAGT GCTACCATGG
 501 GCCAACTGTA TGAGGACAAT CATGAGGAAG ACTATTTTCT GTATGTGGCC
 551 TACAGTGATG AGAGTGTCTA TGGGAAATGA GTGGTTGGAA GCCCAGCAGA
 601 TGGGAGCACC TGGACTTGGG GGTAGGGGAG GGGTGTGTGT GCGCGACATG
 651 GGGAAAGAGG GTGGCTCCCA CCGCAAGGAG ACAGAAGGTG AAGACATCTA
 701 GAAACATTAC ACCACACACA CCGTCATCAC ATTTTCACAT GCTCAATTGA
 751 TATTTTTTGC TGCTTCCTCG GCCCAGGGAG AAAGCATGTC AGGACAGAGC
 801 TGTTGGATTG GCTTTGATAG AGGAATGGGG ATGATGTAAG TTTACAGTAT
 851 TCCTGGGGTT TAATTGTTGT GCAGTTTCAT AGATGGGTCA GGAGGTGGAC
 901 AAGTTGGGGC CAGAGATGAT GGCAGTCCAG CAGCAACTCC CTGTGCTCCC
 951 TTCTCTTTGG GCAGAGATTC TATTTTTGAC ATTTGCACAA GACAGGTAGG
1001 GAAAGGGGAC TTGTGGTAGT GGACCATACC TGGGGACCAA AAGAGACCCA
1051 CTGTAATTGA TGCATTGTGG CCCCTGATCT TCCCTGTCTC ACACTTCTTT
1101 TCTCCCATCC CGGTTGCAAT CTCACTCAGA CATCACAGTA CCACCCCAGG
1151 GGTGGCAGTA GACAACAACC CAGAAATTTA GACAGGGATC TCTTACCTTT
1201 GGAAAATAGG GGTTAGGCAT GAAGGTGGTT GTGATTAAGA AGATGGTTTT
1251 GTTATTAAAT AGCATTAAAC TGGAATTGAC AAGAGTGTTG AGCATCCCTG
1301 TCTAACCTGC TCTTTCTCTT TGGTGCCCCT TATCTCACCC CTTCCTTGGA
1351 ATTTAATAAG TCTCAGGCAT TTCCAATTGT AGACTAAAAC CACTCTTAGC
1401 ATCTCCTCTA GTATTTTCCA TGTATCAGGA AAGAGGTGTC TTATGTAGGG
1451 AGGGGGCAAG TATGAAGTAA GGTAATTATA TACTACTCTC ATTCAGGATT
1501 CTTGCTCCCA TGCTGCTGTC CCTTCAGGCT CACATGCACA GGAATGCTAC
1551 ATGATGGCCA GCTGCTTCCC TCCTTGGTTA TCATCCACTG CAGCTGCTAG
1601 TTAGAAAGGT TTGGAGGGAT GACTTTTAGT AAATCATGGG GATTTTATTG
1651 ATTTATTTTC ACTTTTGGGA TTTTGTGGGG TGGGAGTGGG GAGCAGGAAT
1701 TGCACTCAGA CATGACATTT CAATTCATCT CTGCTAATGA AAAGGGTTCT
1751 TTCTCTTGGG GGAAATGTGT GTGTCAGTTC TGTCAGCTGC AAGTTCTTGT
1801 ATAATGAAGT CAATGCCATC AGGCCAAGGA AATAAAATAA TTGCTTACCT
1851 TAAAAATCGA AAAAAAAAAA AAAAAAAAAC
```

BLAST Results

Entry HS418210 from database EMBL:
human STS SHGC-10496.
Score = 1916, P = 4.0e-80, identities = 394/400

Entry AC006514 from database EMBLNEW:
*** SEQUENCING IN PROGRESS *** Homo sapiens; HTGS phase 1, 68 unordered pieces.
Score = 610, P = 2.7e-16, identities = 128/134
4 exons

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 227 bp to 577 bp; peptide length: 117
Category: strong similarity to known protein

```
  1 MKFQYKEDHP FEYRKKEGEK IRKKYPDRVP VIVEKAPKAR VPDLDKRKYL
 51 VPSDLTVGQF YFLIRKRIHL RPEDALFFFV NNTIPPTSAT MGQLYEDNHE
101 EDYFLYVAYS DESVYGK
```

BLASTP hits

Entry YQD9_CAEEL from database SWISSPROT:
HYPOTHETICAL 14.8 KD PROTEIN C32D5.9 IN CHROMOSOME II.
Score = 496, P = 1.8e-47, identities = 91/116, positives = 105/116

Entry SYRP_LACBI from database SWISSPROT:
SYMBIOSIS-RELATED PROTEIN.
Score = 390, P = 3.1e-36, identities = 68/117, positives = 94/117

Entry LBU93506_1 from database TREMBL:
product: "symbiosis-related protein"; Laccaria bicolor
symbiosis-related protein mRNA, partial cds.
Score = 390, P = 3.1e-36, identities = 68/117, positives = 94/117

Entry GEF2_RAT from database SWISSPROT:
GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2).
Score = 373, P = 2.0e-34, identities = 71/116, positives = 88/116

Alert BLASTP hits for DKFZphfbr2_72n12, frame 2

TREMBLNEW:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete
cds., N = 1, Score = 549, P = 4.7e-53

SWISSPROT:GEF2_HUMAN GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2)., N = 1,
Score = 373, P = 2.1e-34

```
>TREMBLNEW:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete
     cds.
          Length = 117

  HSPs:

 Score = 549 (82.4 bits), Expect = 4.7e-53, P = 4.7e-53
 Identities = 101/116 (87%), Positives = 110/116 (94%)

Query:     1 MKFQYKEDHPFEYRKKEGEKIRKKYPDRVPVIVEKAPKARVPDLDKRKYLVPSDLTVGQF 60
             MKF YKE+HPFE R+ EGEKIRKKYPDRVPVIVEKAPKAR+ DLDK+KYLVPSDLTVGQF
Sbjct:     1 MKFVYKEEHPFEKRRSEGEKIRKKYPDRVPVIVEKAPKARIGDLDKKKYLVPSDLTVGQF 60

Query:    61 YFLIRKRIHLRPEDALFFFVNNTIPPTSATMGQLYEDNHEEDYFLYVAYSDESVYG 116
             YFLIRKRIHLR EDALFFFVNN IPPTSATMGQLY+++HEED+FLY+AYSDESVYG
Sbjct:    61 YFLIRKRIHLRAEDALFFFVNNVIPPTSATMGQLYQEHHEEDFFLYIAYSDESVYG 116
```

Pedant information for DKFZphfbr2_72n12, frame 2

Report for DKFZphfbr2_72n12.2

[LENGTH]	117
[MW]	14044.07
[pI]	8.67
[HOMOL]	TREMBL:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete cds. 1e-56

```
[FUNCAT]       30.03 organization of cytoplasm       [S. cerevisiae, YBL078c] 4e-36
[FUNCAT]       08.22 cytoskeleton-dependent transport       [S. cerevisiae, YBL078c] 4e-36
[FUNCAT]       06.13.04 lysosomal and vacuolar degradation  [S. cerevisiae, YBL078c] 4e-36
[SUPFAM]       hypothetical protein YBL078c 8e-35
[PROSITE]      ASN_GLYCOSYLATION      1
[KW]           Alpha_Beta

SEQ     MKFQYKEDHPFEYRKKEGEKIRKKYPDRVPVIVEKAPKARVPDLDKRKYLVPSDLTVGQF
PRD     ccccccccchhhhhhhhhhhhhhcccccceeeeeccccccccccccceeecccccchhhh

SEQ     YFLIRKRIHLRPEDALFFFVNNTIPPTSATMGQLYEDNHEEDYFLYVAYSDESVYGK
PRD     hhhhhhhhhhcccccceeeeecccccccchhhhhhhhhccccceeeeeeeccccccc
```

Prosite for DKFZphfbr2_72n12.2

```
PS00001       81->85   ASN_GLYCOSYLATION       PDOC00001
```

(No Pfam data available for DKFZphfbr2_72n12.2)

DKFZphfbr2_78c24
----

group: signal transduction

DKFZphfbr2_78c24 encodes a novel 563 amino acid protein with strong similarity to guanylate-binding proteins (GBPs).

GBPs were originally described as proteins that are strongly induced by interferons and are capable of binding to agarose-immobilized guanine nucleotides. hGBP1, the first of two members of this protein family in humans, represents a novel type of GTPase. The novel protein contains an ATP/GTP-binding site motif A (P-loop) and a RGD cell attachment site. It seems to be a new member of the GBP-family and shows a splicing pattern not described previously.

The new protein can find application in modulating/blocking the response of cells to interferons.

strong similarity to guanine nucleotide-binding protein 1/2
but different "splice variant" aa 211-245 of GBP1/2 missing

Sequenced by MediGenomix

Locus: unknown

Insert length: 2952 bp
Poly A stretch at pos. 2927, polyadenylation signal at pos. 2914

```
   1 CAGTTTCATT AGGCTCTGAA GCCATTACAA AGGTTGCTTA ACTTCTAATT
  51 ATTTGATCAC TGAGGAAAAT CCAGAAAGCT ACACAACACT GAAGGGGTGA
 101 AATAAAAGTC CAGCGATCCA GCGAAAGAAA AGAGAAGTGA CAGAAACAAC
 151 TTTACCTGGA CTGAAGATAA AAGCACAGAC AAGAGAACAA TGCCCTGGAC
 201 ATGGCTCCAG AGATCCACAT GACAGGCCCA ATGTGCCTCA TTGAGAACAC
 251 TAATGGGGAA CTGGTGGCGA ATCCAGAAGC TCTGAAAATC CTGTCTGCCA
 301 TTACACAGCC TGTGGTGGTG GTGGCAATTG TGGGCCTCTA CCGCACAGGA
 351 AAATCCTACC TGATGAACAA GCTAGCTGGG AAGAATAAGG GCTTCTCTCT
 401 GGGCTCCACA GTGAAATCTC ACACCAAAGG AATCTGGATG TGGTGTGTGC
 451 CTCACCCCAA AAAGCCAGAA CACACCTTAG TCCTGCTTGA CACTGAGGGC
 501 CTGGGAGATG TAAAGAAGGG TGACAACCAG AATGACTCCT GGATCTTCAC
 551 CCTGGCCGTC CTCCTGAGCA GCACTCTCGT GTACAATAGC ATGGGAACCA
 601 TCAACCAGCA GGCTATGGAC CAACTGTACT ATGTGACAGA GCTGACACAT
 651 CGAATCCGAT CAAAATCCTC ACCTGATGAG AATGAGAATG AGGATTCAGC
 701 TGACTTTGTG AGCTTCTTCC CAGATTTTGT GTGGACACTG AGAGATTTCT
 751 CCCTGGACTT GGAAGCAGAT GGACAACCCC TCACACCAGA TGAGTACCTG
 801 GAGTATTCCC TGAAGCTAAC GCAAGGTAAC AGGAAGCTTG CCCAGCTTGA
 851 GAAACTACAA GATGAAGAGC TGGACCCTGA ATTTGTGCAA CAAGTAGCAG
 901 ACTTCTGTTC CTACATCTTT AGCAATTCCA AAACTAAAAC TCTTTCAGGA
 951 GGCATCAAGG TCAATGGGCC TTGTCTAGAG AGCCTAGTGC TGACCTATAT
1001 CAATGCTATC AGCAGAGGGG ATCTGCCCTG CATGGAGAAC GCAGTCCTGG
1051 CCTTGGCCCA GATAGAGAAC TCAGCCGCAG TGCAAAAGGC TATTGCCCAC
1101 TATGACCAGC AGATGGGCCA GAAGGTGCAG CTGCCCGCAG AAACCCTCCA
1151 GGAGCTGCTG GACCTGCACA GGGTTAGTGA GAGGGAGGCC ACTGAAGTCT
1201 ATATGAAGAA CTCTTTCAAG GATGTGGACC ATCTGTTTCA AAAGAAATTA
1251 GCGGCCCAGC TAGACAAAAA GCGGGATGAC TTTTGTAAAC AGAATCAAGA
1301 AGCATCATCA GATCGTTGCT CAGCTTTACT TCAGGTCATT TTCAGTCCTC
1351 TAGAAGAAGA AGTGAAGGCG GGAATTTATT CGAAACCAGG GGGCTATTGT
1401 CTCTTTATTC AGAAGCTACA AGACCTGGAG AAAAAGTACT ATGAGGAACC
1451 AAGGAAGGGG ATACAGGCTG AAGAGATTCT GCAGACATAC TTGAAATCCA
1501 AGGAGTCTGT GACCGATGCA ATTCTACAGA CAGACCAGAT TCTCACAGAA
1551 AAGGAAAAGG AGATTGAAGT GGAATGTGTA AAAGCTGAAT CTGCACAGGC
1601 TTCAGCAAAA ATGGTGGAGG AAATGCAAAT AAAGTATCAG CAGATGATGG
1651 AAGAGAAAGA GAAGAGTTAT CAAGAACATG TGAAACAATT GACTGAGAAG
1701 ATGGAGAGGG AGAGGGCCCA GTTGCTGGAA GAGCAAGAGA AGACCCTCAC
1751 TAGTAAACTT CAGGAACAGG CCCGAGTACT AAAGGAGAGA TGCCAAGGTG
1801 AAAGTACCCA ACTTCAAAAT GAGATACAAA AGCTACAGAA GACCCTGAAA
1851 AAAAAAACCA AGAGATATAT GTCGCATAAG CTAAAGATCT AAACAACAGA
1901 GCTTTTCTGT CATCCTAACC CAAGGCATAA CTGAAACAAT TTTAGAATTT
1951 GGAACAAGTG TCACTATATT TGATAATAAT TAGATCTTGC ATCATAACAC
2001 TAAAAGTTTA CAAGAACATG CAGTTCAATG ATCAAAATCA TGTTTTTTCC
2051 TTAAAAAGAT TGTAAATTGT GCAACAAAGA TGCATTTACC TCTGTACCAA
2101 CAGAGGAGGG ATCATGAGTT GCCACCACTC AGAAGTTTAT TCTTCCAGAC
2151 GACCAGTGGA TACTGAGGAA AGTCTTAGGT AAAAATCTTG GGACATATTT
2201 GGGCACTGGT TTGGCCAAGT GTACAATAGG TCCCAATATC AGAAACAACC
2251 ATCCTAGCTT CCTAGGGAAG ACAGTGTACA GTTCTCCATT ATATCAAGGC
2301 TACAAGGTCT ATGAGCAATA ATGTGATTTC TGGACATTGC CCATGGATAA
2351 TTCTCACTGA TGGATCTCAA GCTAAAGCAA ACCATCTTAT ACAGAGATCT
2401 AGAATCTTAT ATTTTCCATA GGAAGGTAAA GAAATCATTA GCAAGAGTAG
2451 GAATTGAATC ATAAACAAAT TGGCTAATGA AGAAATCTTT TCTTTCTTGT
2501 TCAATTCATC TAGATTATAA CCTTAATGTG ACACCTGAGA CCTTTAGACA
```

```
2551 GTTGACCCTG AATTAAATAG TCACATGGTA ACAATTATGC ACTGTGTAAT
2601 TTTAGTAATG TATAACATGC AATGATGCAC TTTAACTGAA GATAGAGACT
2651 ATGTTAGAAA ATTGAACTAA TTTAATTATT TGATTGTTTT AATCCTAAAG
2701 CATAAGTTAG TCTTTTCCTG ATTCTTAAAG GTCATACTTG AAATCCTGCC
2751 AATTTTCCCC AAAGGGAATA TGGAATTTTT TTTGACTTTC TTTTGAGCAA
2801 TAAAATAATT GTCTTGCCAT TACTTAGTAT ATGTAGACTT CATCCCAATT
2851 GTCAAACATC CTAGGTAAGT GGTTGACATT TCTTACAGCA ATTACAGATT
2901 ATTTTTGAAC TAGAAATAAA CTAAACTAGA AACAAAAAAA AAAAAAAAAA
2951 AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 201 bp to 1889 bp; peptide length: 563
Category: strong similarity to known protein
Classification: Cell signaling/communication
Prosite motifs: RGD (272-275)
ATP_GTP_A (45-53)

```
  1 MAPEIHMTGP MCLIENTNGE LVANPEALKI LSAITQPVVV VAIVGLYRTG
 51 KSYLMNKLAG KNKGFSLGST VKSHTKGIWM WCVPHPKKPE HTLVLLDTEG
101 LGDVKKGDNQ NDSWIFTLAV LLSSTLVYNS MGTINQQAMD QLYYVTELTH
151 RIRSKSSPDE NENEDSADFV SFFPDFVWTL RDFSLDLEAD GQPLTPDEYL
201 EYSLKLTQGN RKLAQLEKLQ DEELDPEFVQ QVADFCSYIF SNSKTKTLSG
251 GIKVNGPCLE SLVLTYINAI SRGDLPCMEN AVLALAQIEN SAAVQKAIAH
301 YDQQMGQKVQ LPAETLQELL DLHRVSEREA TEVYMKNSFK DVDHLFQKKL
351 AAQLDKKRDD FCKQNQEASS DRCSALLQVI FSPLEEEVKA GIYSKPGGYC
401 LFIQKLQDLE KKYYEEPRKG IQAEEILQTY LKSKESVTDA ILQTDQILTE
451 KEKEIEVECV KAESAQASAK MVEEMQIKYQ QMMEEKEKSY QEHVKQLTEK
501 MERERAQLLE EQEKTLTSKL QEQARVLKER CQGESTQLQN EIQKLQKTLK
551 KKTKRYMSHK LKI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78c24, frame 3

PIR:A41268 guanine nucleotide-binding protein 1 - human, N = 2, Score = 1306, P = 4.9e-238

PIR:A46459 macrophage-activation gene-1 protein mag-1 - mouse, N = 2, Score = 942, P = 8.9e-184

PIR:S70524 guanine nucleotide-binding protein 2 - human, N = 2, Score = 1131, P = 4.1e-210

TREMBL:AF077007_1 gene: "Gbp2"; product: "interferon-induced guanylate binding protein GBP-2"; Mus musculus interferon-induced guanylate binding protein GBP-2 (Gbp2) mRNA, complete cds., N = 2, Score = 904, P = 1.2e-179

```
>PIR:A41268 guanine nucleotide-binding protein 1 - human
           Length = 592

  HSPs:

 Score = 1306 (195.9 bits), Expect = 4.9e-238, Sum P(2) = 4.9e-238
 Identities = 264/332 (79%), Positives = 288/332 (86%)

Query:   211 RKLAQLEKLQDEELDPEFVQQVADFCSYIFSNSKTKTLSGGIKVNGPCLESLVLTYINAI 270
             RKLAQLEKLQDEELDPEFVQQVADFCSYIFSNSKTKTLSGGI+VNGP LESLVLTY+NAI
Sbjct:   245 RKLAQLEKLQDEELDPEFVQQVADFCSYIFSNSKTKTLSGGIQVNGPRLESLVLTYVNAI 304
```

```
Query:   271 SRGDLPCMENAVLALAQIENSAAVQKAIAHYDQQMGQKVQLPAETLQELLDLHRVSEREA 330
             S GDLPCMENAVLALAQIENSAAVQKAIAHY+QQMGQKVQLP E+LQELLDLHR SEREA
Sbjct:   305 SSGDLPCMENAVLALAQIENSAAVQKAIAHYEQQMGQKVQLPTESLQELLDLHRDSEREA 364

Query:   331 TEVYMKNSFKDVDHLFQKKLAAQLDKKRDDFCKQNQEASSDRCSALLQVIFSPLEEEVKA 390
              EV++++SFKDVDHLFQK+LAAQL+KKRDDFCKQNQEASSDRCS LLQVIFSPLEEEVKA
Sbjct:   365 IEVFIRSSFKDVDHLFQKELAAQLEKKRDDFCKQNQEASSDRCSGLLQVIFSPLEEEVKA 424

Query:   391 GIYSKPGGYCLFIQKLQDLEKKYYEEPRKGIQAEEILQTYLKSKESVTDAILQTDQILTX 450
             GIYSKPGGY LF+QKLQDL+KKYYEEPRKGIQAEEILQTYLKSKES+TDAILQTDQ LT
Sbjct:   425 GIYSKPGGYRLFVQKLQDLKKKYYEEPRKGIQAEEILQTYLKSKESMTDAILQTDQTLTE 484

Query:   451 XXXXXXXXXXXXXSAQASAKMVEEMQIKYQQMMEEKEKSYQEHVKQLTEKMXXXXXXXXX 510
                          SAQASAKM++EMQ K +QMME+KE+SYQEH+KQLTEKM
Sbjct:   485 KEKEIEVERVKAESAQASAKMLQEMQRKNEQMMEQKERSYQEHLKQLTEKMENDRVQLLK 544

Query:   511 XXXKTLTSKLQEQARVLKERCQGESTQLQNEI 542
                +TL  KLQEQ ++LKE  Q ES  ++NEI
Sbjct:   545 EQERTLALKLQEQEQLLKEGFQKESRIMKNEI 576

 Score = 1012 (151.8 bits), Expect = 4.9e-238, Sum P(2) = 4.9e-238
 Identities = 194/211 (91%), Positives = 200/211 (94%)

Query:     1 MAPEIHMTGPMCLIENTNGELVANPEALKILSAITQPVVVVAIVGLYRTGKSYLMNKLAG 60
             MA EIHMTGPMCLIENTNG L+ANPEALKILSAITQP+VVVAIVGLYRTGKSYLMNKLAG
Sbjct:     1 MASEIHMTGPMCLIENTNGRLMANPEALKILSAITQPMVVVAIVGLYRTGKSYLMNKLAG 60

Query:    61 KNKGFSLGSTVKSHTKGIWMWCVPHPKKPEHTLVLLDTEGLGDVKKGDNQNDSWIFTLAV 120
             K KGFSLGSTV+SHTKGIWMWCVPHPKKP H LVLLDTEGLGDV+KGDNQNDSWIF LAV
Sbjct:    61 KKKGFSLGSTVQSHTKGIWMWCVPHPKKPGHILVLLDTEGLGDVEKGDNQNDSWIFALAV 120

Query:   121 LLSSTLVYNSMGTINQQAMDQLYYVTELTHRIRSKSSPDENENE--DSADFVSFFPDFVW 178
             LLSST VYNS+GTINQQAMDQLYYVTELTHRIRSKSSPDENENE  DSADFVSFFPDFVW
Sbjct:   121 LLSSTFVYNSIGTINQQAMDQLYYVTELTHRIRSKSSPDENENEVEDSADFVSFFPDFVW 180

Query:   179 TLRDFSLDLEADGQPLTPDEYLEYSLKLTQG 209
             TLRDFSLDLEADGQPLTPDEYL YSLKL +G
Sbjct:   181 TLRDFSLDLEADGQPLTPDEYLTYSLKLKKG 211
```

Pedant information for DKFZphfbr2_78c24, frame 3

Report for DKFZphfbr2_78c24.3

```
[LENGTH]     563
[MW]         64127.72
[pI]         5.45
[HOMOL]      PIR:A41268 guanine nucleotide-binding protein 1 - human 0.0
[SUPFAM]     guanine nucleotide-binding protein 1 0.0
[PROSITE]    ATP_GTP_A      1
[PROSITE]    RGD    1
[KW]         TRANSMEMBRANE  1
[KW]         LOW_COMPLEXITY      6.75 %
[KW]         COILED_COIL      10.48 %

SEQ    MAPEIHMTGPMCLIENTNGELVANPEALKILSAITQPVVVVAIVGLYRTGKSYLMNKLAG
SEG    ............................................................
PRD    cccccccccceeeeecccccchhhhhhhhhhhhhhcceeeeeeeeeccccccchhhhhhhhh
COILS  ............................................................
MEM    ...........................MMMMMMMMMMMMMMMMM..............

SEQ    KNKGFSLGSTVKSHTKGIWMWCVPHPKKPEHTLVLLDTEGLGDVKKGDNQNDSWIFTLAV
SEG    ............................................................
PRD    cccccccccccccccceeeeeeccccccccceeeeeeecccccccccccccccchhhhhhhh
COILS  ............................................................
MEM    ............................................................

SEQ    LLSSTLVYNSMGTINQQAMDQLYYVTELTHRIRSKSSPDENENEDSADFVSFFPDFVWTL
SEG    ............................................................
PRD    hhhhheeecccccchhhhhhhhhhhhhhhhhhhhhhhcccccccccccccceeeecccceeeeh
COILS  ............................................................
MEM    ............................................................

SEQ    RDFSLDLEADGQPLTPDEYLEYSLKLTQGNRKLAQLEKLQDEELDPEFVQQVADFCSYIF
SEG    ............................................................
PRD    hhhhhhhhcccccccchhhhhhhhhhcchhhhhhhhhhhhhccccchhhhhhhhhhhhc
COILS  ............................................................
```

```
MEM     ............................................................

SEQ     SNSKTKTLSGGIKVNGPCLESLVLTYINAISRGDLPCMENAVLALAQIENSAAVQKAIAH
SEG     ............................................................
PRD     cccceeeccccccccccchhhhhhhhhhcccccccchhhhhhhhhhhhhhhhhhhhhhhh
COILS   ............................................................
MEM     ............................................................

SEQ     YDQQMGQKVQLPAETLQELLDLHRVSEREATEVYMKNSFKDVDHLFQKKLAAQLDKKRDD
SEG     ............................................................
PRD     hhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhh
COILS   ............................................................
MEM     ............................................................

SEQ     FCKQNQEASSDRCSALLQVIFSPLEEEVKAGIYSKPGGYCLFIQKLQDLEKKYYEEPRKG
SEG     ............................................................
PRD     hhhhhhchhhhhhhhhhhhhhhhhhhccccccccccccceeehhhhhhhhhhhhhccccc
COILS   ............................................................
MEM     ............................................................

SEQ     IQAEEILQTYLKSKESVTDAILQTDQILTEKEKEIEVECVKAESAQASAKMVEEMQIKYQ
SEG     .............................xxxxxxxxxxxxxx.................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ............................................................
MEM     ............................................................

SEQ     QMMEEKEKSYQEHVKQLTEKMERERAQLLEEQEKTLTSKLQEQARVLKERCQGESTQLQN
SEG     .....................xxxxxxxxxxxx...........................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhh
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC........CCCCCCCCCCCCCCCCCCCCCC
MEM     ............................................................

SEQ     EIQKLQKTLKKKTKRYMSHKLKI
SEG     ..xxxxxxxxxxxx.........
PRD     hhhhhhhhhhhhhhhhhhhhccc
COILS   CCCCCCC................
MEM     .......................
```

Prosite for DKFZphfbr2_78c24.3

PS00016	272->275	RGD	PDOC00016
PS00017	45->53	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphfbr2_78c24.3)

DKFZphfbr2_78d13
----

group: brain derived

DKFZphfbr2_78d13 encodes a novel 259 amino acid protein with similarity to C. elegans putative protein from cosmid K08B12.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to C.elegans K08B12.3

Sequenced by MediGenomix

Locus: /map="338.4 cR from top of Chr18 linkage group"

Insert length: 2195 bp
Poly A stretch at pos. 2175, polyadenylation signal at pos. 2156

```
   1 CGTCCGTCGG GCAGCAGCGG GGCTGTCTAT CCCGGCTGAG GACCCGCGGC
  51 CAGTGCGGGT GGCTGGCTTT GCCATTAGCG GGGGCCTTTC CTGAGGACGG
 101 CGTACGGAGT GTGGGGAATG AAGGATGGCA GCATGCCGTG CATTAAAAGC
 151 TGTTTTGGTA GATCTCAGTG GCACACTTCA CATTGAAGAT GCAGCTGTGC
 201 CAGGCGCACA GGAAGCTCTT AAAAGGTTAC GTGGTGCTTC TGTAATCATT
 251 AGGTTTGTGA CCAATACAAC CAAAGAGAGC AAGCAAGACC TGTTAGAAAG
 301 GTTGAGAAAA TTGGAATTTG ATATCTCTGA AGATGAAATA TTCACATCTC
 351 TGACTGCAGC CAGAAGTTTA CTAGAGCGGA AACAAGTCAG ACCCATGCTG
 401 CTAGTTGATG ATCGGGCACT ACCTGATTTC AAAGGAATAC AAACAAGTGA
 451 TCCTAATGCT GTGGTCATGG GATTGGCACC AGAACATTTT CATTATCAAA
 501 TTCTGAATCA AGCATTCCGG TTACTCCTGG ATGGAGCACC TCTGATAGCA
 551 ATCCACAAAG CCAGGTATTA CAAGAGGAAA GATGGCTTAG CCCTGGGGCC
 601 TGGACCATTT GTGACTGCTT TAGAGTATGC CACAGATACC AAAGCCACAG
 651 TCGTGGGGAA ACCAGAGAAG ACGTTCTTTT TGGAAGCATT GCGGGGCACT
 701 GGCTGTGAAC CTGAGGAGGC TGTCATGATA GGAGATGATT GCAGGGATGA
 751 TGTTGGTGGG GCTCAAGATG TCGGCATGCT GGGCATCTTA GTAAAGACTG
 801 GGAAATATCG AGCATCAGAT GAAGAAAAAA TTAATCCACC TCCTTACTTA
 851 ACTTGTGAGA GTTTCCCTCA TGCTGTGGAC CACATTCTGC AGCACCTATT
 901 GTGAAGCAAT GTGTGCATCT GAAGCAACTT GAAATGCAGC TTCTTATTGT
 951 CTGGAATGAA TCCCTTACCA ACTCAGTGCC AGCATCGGTA GACACCAGTC
1001 AGTGCTGATC GCTTTTTAAC CCTCTTTTGT TGTGCATTAA TTAGAAAGAA
1051 AGGTATTGAA TTGCGGCTAG CCAGTAAGCC TTGCTAATCT CTTTTATTTT
1101 GTAACTGAAG ATGAGACCCA AAGAAAGGGA AAGCTGAGAT TTTGTGCCAT
1151 TCCTTTTAAA ATATTCATCA GGTTAGGTGG GCCTGTGGGG GAAAAGCTAC
1201 TACAGGGAAG AGTGTTCTCT GCTGTCTCTT CACTGGAAAA CAGGGAGGGG
1251 GGATTTCAGA CTGTGAAGAA AGTTGAATGG TGGTTTTTAA ATTATAAAGT
1301 AATGTATTAA AAGGTGCATT AGGCTGTAGT TCTAATATTG AGTTCAACTG
1351 TGAAATCCAT CAGATGTGCC AAATGGAGAA GACAGAAAGC AACAAAGTGA
1401 ATTGTTCTTT AGCCCAAGTG GTACAGTGAA TTTGCTTTAA CAGATGTTGA
1451 AAACTAAATT TTCTACTGTA TTCCCAGCAC GGGTGACTTC TTTTTCTCTT
1501 CATTAGCCAG AGATGACTAA TTTAAATTTA GAACCAGATT TTAATTTAAA
1551 TTAATATTTC CATTAATAAC CTACTCATTG CAGATACCTA TTATACTGTG
1601 TAACAGTTGT TTTGGAAATT TTATGTAAAA TTAAAACTAT CAGTATTTTA
1651 CAGATGTTTT AATTAGACAT TGTTATTAAC AGGAACAGTG CAGAAACTAG
1701 AATCAAGCCT TATAATATCT TATAGACCAT GCATTTTTGA AGTTAGTGTC
1751 CACTAGGGTC CTATTAACTG TACATTTGCA AGATTTCATT ATTTTTGCCT
1801 CTGACACTAT GGGAAAAATT TTTTAGAAGC TATTGGGACA GATTCAAGCT
1851 TTTATGCACT TGGTTACTAC AGCTGTAAAA TGAAATCTCG TCTTGTAGCA
1901 TGGATTATTC TTCTCATGTT AAACCCACCA AAATAAAGGG GACTAAATAG
1951 GTAATGATTT TCCTAGTGCA TTTGCATACT GTGATAATCC TGGGCCTTGC
2001 AATAGTTCTA CAGGGCTCTT GGGCATTGAA TTATTAGGAT GTAATTGTAC
2051 ATCATTGTAG TGTTCACCTT ATTGAAGCTC ACTCTGATGT TAATGAGCTT
2101 CGGGTTTTGA TGCTTGTTTA GAGATCAGCA GTCTTGGATG GGAGGGAACA
2151 AAGCTAAATA AATGTTAGTT TGGTGAAAAA AAAAAAAAAA AAAAA
```

BLAST Results

Entry HS599355 from database EMBL:
human STS WI-13484.
Score = 1262, P = 3.6e-52, identities = 274/289

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 125 bp to 901 bp; peptide length: 259
Category: similarity to unknown protein
Classification: no clue

```
  1 MAACRALKAV LVDLSGTLHI EDAAVPGAQE ALKRLRGASV IIRFVTNTTK
 51 ESKQDLLERL RKLEFDISED EIFTSLTAAR SLLERKQVRP MLLVDDRALP
101 DFKGIQTSDP NAVVMGLAPE HFHYQILNQA FRLLLDGAPL IAIHKARYYK
151 RKDGLALGPG PFVTALEYAT DTKATVVGKP EKTFFLEALR GTGCEPEEAV
201 MIGDDCRDDV GGAQDVGMLG ILVKTGKYRA SDEEKINPPP YLTCESFPHA
251 VDHILQHLL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78d13, frame 2

TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid K08B12., N = 1, Score = 609, P = 2.2e-59

TREMBL:CEC13C4_5 gene: "C13C4.4"; Caenorhabditis elegans cosmid C13C4, N = 1, Score = 408, P = 4.4e-38

```
>TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid
     K08B12.
          Length = 257

  HSPs:

 Score = 609 (91.4 bits), Expect = 2.2e-59, P = 2.2e-59
 Identities = 132/251 (52%), Positives = 172/251 (68%)

Query:     7 LKAVLVDLSGTLHIEDAAVPGAQEALKRLRGASVIIRFVTNTTKESKQDLLERLRKLEFD 66
               + +VL+DLSGT+HIE+ A+PGAQ AL+ LR  + + +FVTNTTKESK+ L +RL    F
Sbjct:     4 ISSVLIDLSGTIHIEEFAIPGAQTALELLRQHAKV-KFVTNTTKESKRLLHQRLINCGFK 62

Query:    67 ISEDEIFTSLTAARSLLERKQVRPMLLVDDRALPDFKGIQTSDPNAVVMGLAPEHFHYQI 126
               + ++EIFTSLTAAR L+ + Q RP  +VDDRA+ DF+GI T DPNAVV+GLAPE F+
Sbjct:    63 VEKEEIFTSLTAARDLIVKNQYRPFFIVDDRAMEDFEGISTDDPNAVVIGLAPEKFNDTT 122

Query:   127 LNQAFRLLLDG-APLIAIHKARYYKRKDGLALGPGPFVTALEYATDTKATVVGKPEKTFF 185
             L  AFRL+ +  A LIAI+K RY++   GL LGPG +V  LEY+   +AT+VGKP K FF
Sbjct:   123 LTHAFRLIKEKKASLIAINKGRYHQTNAGLCLGPGTYVAGLEYSAGVEATIVGKPNKLFF 182

Query:   186 LEALRGTG--CEPEEAVMIGDDCRDDVGGAQDVGMLGILVKTGKYRASDEEKINPPPYLT 243
               AL+       +  AVMIGDD  DD  GA  +GM  ILVKTGK+R  DE K+
Sbjct:   183 ESALQSLNENVDFSSAVMIGDDVNDDALGAIKIGMRAILVKTGKFRDGDELKVKN----V 238

Query:   244 CESFPHAVDHILQH 257
                SF  AV+ I+++
Sbjct:   239 ANSFVDAVNMIIEN 252
```

Pedant information for DKFZphfbr2_78d13, frame 2

Report for DKFZphfbr2_78d13.2

[LENGTH]	259
[MW]	28536.04
[pI]	5.84
[HOMOL]	TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid K08B12. 3e-62
[FUNCAT]	r general function prediction [M. jannaschii, MJ1437] 3e-05
[SUPFAM]	nagD protein 4e-18
[KW]	Alpha_Beta

```
SEQ    MAACRALKAVLVDLSGTLHIEDAAVPGAQEALKRLRGASVIIRFVTNTTKESKQDLLERL
PRD    cccccccceeeeeecccceeeecccccchhhhhhhhhhccceeeeeeccccchhhhhhhhh

SEQ    RKLEFDISEDEIFTSLTAARSLLERKQVRPMLLVDDRALPDFKGIQTSDPNAVVMGLAPE
PRD    hhhcccccccceeeehhhhhhhhhhhhcccceeeeeechhhhhhcccccccccceeeeecccc

SEQ    HFHYQILNQAFRLLLDGAPLIAIHKARYYKRKDGLALGPGPFVTALEYATDTKATVVGKP
PRD    chhhhhhhhhhhhhhhcccceeeeecccccccccccccccccchhhhhhhhcccceeeeccc

SEQ    EKTFFLEALRGTGCEPEEAVMIGDDCRDDVGGAQDVGMLGILVKTGKYRASDEEKINPPP
PRD    cchhhhhhhhhhcccccceeeeecccchhhhhhhhhccceeeeeeeccccccccccccccc

SEQ    YLTCESFPHAVDHILQHLL
PRD    ccccccchhhhhhhhhhccc
```

(No Prosite data available for DKFZphfbr2_78d13.2)

(No Pfam data available for DKFZphfbr2_78d13.2)

DKFZphfbr2_78k24
----

group: metabolism

DKFZphfbr2_78k24 encodes a novel 372 amino acid protein with similarity to Mus musculus ubiquitin specific protease UBP43.

The novel protein contains a Prosite ubiquitin carboxyl-terminal hydrolases family 2 signature 2. Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquinated proteins.

The new protein can find application in modulation of protein stability/degradation in cells.

Ubiquitin carboxyl-terminal hydrolases family 2 signature 2.

strong similarity to mouse ubiquitin specific protease UBP43

Sequenced by MediGenomix

Locus: unknown

Insert length: 1874 bp
Poly A stretch at pos. 1852, polyadenylation signal at pos. 1836

```
   1 AGTCCCGACG TGGAACTCAG CAGCGGAGGC TGGACGCTTG CATGGCGCTT
  51 GAGAGATTCC ATCGTGCCTG GCTCACATAA GCGCTTCCTG GAAGTGAAGT
 101 CGTGCTGTCC TGAACGCGGG CCAGGCAGCT GCGGCCTGGG GGTTTTGGAG
 151 TGATCACGAA TGAGCAAGGC GTTTGGGCTC CTGAGGCAAA TCTGTCAGTC
 201 CATCCTGGCT GAGTCCTCGC AGTCCCCGGC AGATCTTGAA GAAAAGAAGG
 251 AAGAAGACAG CAACATGAAG AGAGAGCAGC CCAGAGAGCG TCCCAGGGCC
 301 TGGGACTACC CTCATGGCCT GGTTGGTTTA CACAACATTG GACAGACCTG
 351 CTGCCTTAAC TCCTTGATTC AGGTGTTCGT AATGAATGTG GACTTCACCA
 401 GGATATTGAA GAGGATCACG GTGCCCAGGG GAGCTGACGA GCAGAGGAGA
 451 AGCGTCCCTT TCCAGATGCT TCTGCTGCTG GAGAAGATGC AGGACAGCCG
 501 GCAGAAAGCA GTGCGGCCCC TGGAGCTGGC CTACTGCCTG CAGAAGTGCA
 551 ACGTGCCCTT GTTTGTCCAA CATGATGCTG CCCAACTGTA CCTCAAACTC
 601 TGGAACCTGA TTAAGGACCA GATCACTGAT GTGCACTTGG TGGAGAGACT
 651 GCAGGCCCTG TATACGATCC GGGTGAAGGA CTCCTTGATT TGCGTTGACT
 701 GTGCCATGGA GAGTAGCAGA AACAGCAGCA TGCTCACCCT CCCACTTTCT
 751 CTTTTTGATG TGGACTCAAA GCCCCTGAAG ACACTGGAGG ACGCCCTGCA
 801 CTGCTTCTTC CAGCCCAGGG AGTTATCAAG CAAAAGCAAG TGCTTCTGTG
 851 AGAACTGTGG GAAGAAGACC CGTGGGAAAC AGGTCTTGAA GCTGACCCAT
 901 TTGCCCCAGA CCCTGACAAT CCACCTCATG CGATTCTCCA TCAGGAATTC
 951 ACAGACGAGA AAGATCTGCC ACTCCCTGTA CTTCCCCCAG AGCTTGGATT
1001 TCAGCCAGAT CCTTCCAATG AAGCGAGAGT CTTGTGATGC TGAGGAGCAG
1051 TCTGGAGGGC AGTATGAGCT TTTTGCTGTG ATTGCGCACG TGGGAATGGC
1101 AGACTCCGGT CATTACTGTG TCTACATCCG GAATGCTGTG GATGGAAAAT
1151 GGTTCTGCTT CAATGACTCC AATATTTGCT TGGTGTCCTG GGAAGACATC
1201 CAGTGTACCT ACGGAAATCC TAACTACCAC TGGCAGGAAA CTGCATATCT
1251 TCTGGTTTAC ATGAAGATGG AGTGCTAATG GAAATGCCCA AAACCTTCAG
1301 AGATTGACAC GCTGTCATTT TCCATTTCCG TTCCTGGATC TACGGAGTCT
1351 TCTAAGAGAT TTTGCAATGA GGAGAAGCAT TGTTTTCAAA CTATATAACT
1401 GAGCCTTATT TATAATTAGG GATATTATCA AAATATGTAA CCATGAGGCC
1451 CCTCAGGTCC TGATCAGTCA GAATGGATGC TTTCACCAGC AGACCCGGCC
1501 ATGTGGCTGC TCGGTCCTGG GTGCTCGCTG CTGTGCAAGA CATTAGCCCT
1551 TTAGTTATGA GCCTGTGGGA ACTTCAGGGG TTCCCAGTGG GGAGAGCAGT
1601 GGCAGTGGGA GGCATCTGGG GGCCAAAGGT CAGTGGCAGG GGGTATTTCA
1651 GTATTATACA ACTGCTGTGA CCAGACTTGT ATACTGGCTG AATATCAGTG
1701 CTGTTTGTAA TTTTTCACTT TGAGAACCAA CATTAATTCC ATATGAATCA
1751 AGTGTTTTGT AACTGCTATT CATTTATTCA GCAAATATTT ATTGATCATC
1801 TCTTCTCCAT AAGATAGTGT GATAAACACA GTCATGAATA AAGTTATTTT
1851 CCACAAAAAA AAAAAAAAAA AAAA
```

BLAST Results

Entry AC005500 from database EMBL:
, complete sequence.
Score = 859, P = 5.7e-143, identities = 175/179
8 exons matching Bp 317-1230

```
                         Medline entries
                         ---------------

99182491:
A novel ubiquitin-specific protease, UBP43, cloned from leukemia
fusion protein AML1-ETO-expressing mice, functions in
hematopoietic cell differentiation.


                  Peptide information for frame 1
                  -------------------------------

ORF from 160 bp to 1275 bp; peptide length: 372
Category: strong similarity to known protein
Classification: Protein management
Prosite motifs: UCH_2_2 (302-320)

    1 MSKAFGLLRQ ICQSILAESS QSPADLEEKK EEDSNMKREQ PRERPRAWDY
   51 PHGLVGLHNI GQTCCLNSLI QVFVMNVDFT RILKRITVPR GADEQRRSVP
  101 FQMLLLLEKM QDSRQKAVRP LELAYCLQKC NVPLFVQHDA AQLYLKLWNL
  151 IKDQITDVHL VERLQALYTI RVKDSLICVD CAMESSRNSS MLTLPLSLFD
  201 VDSKPLKTLE DALHCFFQPR ELSSKSKCFC ENCGKKTRGK QVLKLTHLPQ
  251 TLTIHLMRFS IRNSQTRKIC HSLYFPQSLD FSQILPMKRE SCDAEEQSGG
  301 QYELFAVIAH VGMADSGHYC VYIRNAVDGK WFCFNDSNIC LVSWEDIQCT
  351 YGNPNYHWQE TAYLLVYMKM EC


                          BLASTP hits

No BLASTP hits available

           Alert BLASTP hits for DKFZphfbr2_78k24, frame 1

TREMBLNEW:AF069502_1 product: "ubiquitin specific protease UBP43"; Mus
musculus ubiquitin specific protease UBP43 mRNA, complete cds., N = 1,
Score = 1367, P = 1e-139

SWISSPROT:UBPE_DROME UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 64E (EC
3.1.2.15) (UBIQUITIN THIOLESTERASE 64E) (UBIQUITIN-SPECIFIC PROCESSING
PROTEASE 64E) (DEUBIQUITINATING ENZYME 64E)., N = 2, Score = 248, P =
5.3e-33


>TREMBLNEW:AF069502_1 product: "ubiquitin specific protease UBP43"; Mus
     musculus ubiquitin specific protease UBP43 mRNA, complete cds.
            Length = 368

  HSPs:

 Score = 1367 (205.1 bits), Expect = 1.0e-139, P = 1.0e-139
 Identities = 262/369 (71%), Positives = 295/369 (79%)

Query:      1 MSKAFGLLRQICQSILAESSQSPADLEEKKEEDSNMKREQPRERPRAWDYPHGLVGLHNI 60
              M K FGLLR+ CQS++AE  Q  A LEE  E     KR   R+   AWD PHGLVGLHNI
Sbjct:      1 MGKGFGLLRKPCQSVVAEPQQYSA-LEE--ERTMKRKRVLSRDLCSAWDSPHGLVGLHNI 57

Query:     61 GQTCCLNSLIQVFVMNVDFTRILKRITVPRGADEQRRSVPFQMLLLLEKMQDSRQKAVRP 120
              GQTCCLNSL+QVF+MN+DF  ILKRITVPR A+E++RSVPFQ+LLLLEKMQDSRQKA+ P
Sbjct:     58 GQTCCLNSLLQVFMMNMDFRMILKRITVPRSAEERKRSVPFQLLLLLEKMQDSRQKALLP 117

Query:    121 LELAYCLQKCNVPLFVQHDAAQLYLKLWNLIKDQITDVHLVERLQALYTIRVKDSLICVD 180
               EL  CLQK NVPLFVQHDAAQLYL +WNL KDQITD  L ERLQ L+TI  ++SLICV
Sbjct:    118 TELVQCLQKYNVPLFVQHDAAQLYLTIWNLTKDQITDTDLTERLQGLFTIWTQESLICVG 177

Query:    181 CAMESSRNSSMLTLPLSLFDVDSKPLKTLEDALHCFFQPRELSSKSKCFCENCGKKTRGK 240
              C  ESSR S +LTL L LFD D+KPLKTLEDAL CF QP+EL+S   C CE CG+KT  K
Sbjct:    178 CTAESSRRSKLLTLSLPLFDKDAKPLKTLEDALRCFVQPKELASSDMC-CETCGEKTPWK 236

Query:    241 QVLKLTHLPQTLTIHLMRFSIRNSQTRKICHSLYFPQSLDFSQILPMKRESCDAEEQSGG 300
              QVLKLTHLPQTLTIHLMRFS RNS+T KICHS+ FPQSLDFSQ+LP  +  D +EQS
Sbjct:    237 QVLKLTHLPQTLTIHLMRFSARNSRTEKICHSVNFPQSLDFSQVLPTEEDLGDTKEQSEI 296

Query:    301 QYELFAVIAHVGMADSGHYCVYIRNAVDGKWFCFNDSNICLVSWEDIQCTYGNPNYHWQE 360
               YELFAVIAHVGMAD GHYC YIRN VDGKWFCFNDS++C V+W+D+QCTYGN  Y W+E
Sbjct:    297 HYELFAVIAHVGMADFGHYCAYIRNPVDGKWFCFNDSHVCWVTWKDVQCTYGNHRYRWRE 356

Query:    361 TAYLLVYMK 369
```

```
         TAYLLVY K
Sbjct:   357 TAYLLVYTK 365
```

Pedant information for DKFZphfbr2_78k24, frame 1

Report for DKFZphfbr2_78k24.1

```
[LENGTH]       372
[MW]           43011.12
[pI]           8.05
[HOMOL]        TREMBLNEW:AF069502_1 product: "ubiquitin specific protease UBP43"; Mus musculus
ubiquitin specific protease UBP43 mRNA, complete cds. 1e-151
[FUNCAT]       06.13 proteolysis      [S. cerevisiae, YMR304w] 3e-19
[FUNCAT]       06.13.01 cytoplasmic degradation      [S. cerevisiae, YJL197w] 3e-16
[FUNCAT]       06.07 protein modification (glycolsylation, acylation, myristylation,
palmitylation, farnesylation and processing)      [S. cerevisiae, YMR223w] 1e-15
[FUNCAT]       04.05.01.04 transcriptional control   [S. cerevisiae, YNL186w] 6e-12
[FUNCAT]       03.10 sporulation and germination     [S. cerevisiae, YDR069c] 9e-11
[FUNCAT]       10.03.99 other osmosensing activities      [S. cerevisiae, YDR069c] 9e-11
[FUNCAT]       30.10 nuclear organization     [S. cerevisiae, YDR069c] 9e-11
[FUNCAT]       30.03 organization of cytoplasm      [S. cerevisiae, YDR069c] 9e-11
[FUNCAT]       09.25 vacuolar and lysosomal biogenesis      [S. cerevisiae, YDR069c] 9e-11
[BLOCKS]       BL00582A Ribosomal protein L33 proteins
[BLOCKS]       BL00972E
[BLOCKS]       BL00972D
[BLOCKS]       BL00972A
[EC]           2.4.2.29 Queuine tRNA-ribosyltransferase 1e-06
[PIRKW]        pentosyltransferase 1e-06
[PIRKW]        glycosyltransferase 1e-06
[PIRKW]        tRNA modification 1e-06
[PIRKW]        alternative splicing 7e-11
[PIRKW]        hydrolase 7e-06
[SUPFAM]       deubiquinating enzyme SSV7 2e-09
[PROSITE]      UCH_2_2 1
[PFAM]         Ubiquitin carboxyl-terminal hydrolases family 2
[PFAM]         Ubiquitin carboxyl-terminal hydrolases family 2
[KW]           Alpha_Beta

SEQ     MSKAFGLLRQICQSILAESSQSPADLEEKKEEDSNMKREQPRERPRAWDYPHGLVGLHNI
PRD     ccccceeechhhhhhhhcccccccchhhhhhhhcccccccccccccccccccccccccccc

SEQ     GQTCCLNSLIQVFVMNVDFTRILKRITVPRGADEQRRSVPFQMLLLLEKMQDSRQKAVRP
PRD     cceeehhhhhhhhhhccccchhhhhhhhcccccccccchhhhhhhhhhhhhhhhhhhccccc

SEQ     LELAYCLQKCNVPLFVQHDAAQLYLKLWNLIKDQITDVHLVERLQALYTIRVKDSLICVD
PRD     hhhhhccccccccchhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhheeeee

SEQ     CAMESSRNSSMLTLPLSLFDVDSKPLKTLEDALHCFFQPRELSSKSKCFCENCGKKTRGK
PRD     ccccccccccccccccccccccccchhhhhhhhhhhhhhcccccccceeecccccccccc

SEQ     QVLKLTHLPQTLTIHLMRFSIRNSQTRKICHSLYFPQSLDFSQILPMKRESCDAEEQSGG
PRD     cceeeecccchhhhhhhhhhhcchhhhhcccccccccccccccccccccccccccccccc

SEQ     QYELFAVIAHVGMADSGHYCVYIRNAVDGKWFCFNDSNICLVSWEDIQCTYGNPNYHWQE
PRD     eeeeeeeeeeccccccceeeeeeccccccceeeecccceeeeeeccccccccccccccchhh

SEQ     TAYLLVYMKMEC
PRD     hhhhhhhhhccc
```

Prosite for DKFZphfbr2_78k24.1

```
PS00973     302->320    UCH_2_2              PDOC00750
```

Pfam for DKFZphfbr2_78k24.1

```
HMM_NAME       Ubiquitin carboxyl-terminal hydrolases family 2

HMM                 *GIqNlGNTCYMNSIIQCL*
                     G+ N+G TC +NS+IQ+
Query          56   GLHNIGQTCCLNSLIQVF     73
```

```
HMM_NAME        Ubiquitin carboxyl-terminal hydrolases family 2

HMM                 *YdLYgVICHYGntldyGHYWaYVKNenhHRWkWYYFDDEtV*
                     Y+L++VI H G   D+GHY +Y++N     ++KW++F+D+++
Query           302  YELFAVIAHVG-MADSGHYCVYIRNAV--DGKWFCFNDSNI     339
```

DKFZphfbr2_78n23
----

group: brain derived

DKFZphfbr2_78n23 encodes a novel 329 amino acid protein with similarity to A.thaliana F26P21.80 protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A.thaliana F26P21.80

Sequenced by MediGenomix

Locus: /map="89.1 cR from top of Chr19 linkage group"

Insert length: 1447 bp
Poly A stretch at pos. 1374, polyadenylation signal at pos. 1353

```
   1 TACAACTTCC GGCTGTAAAG ATGGCGGCTT CCTAGTGAGT CGGCGGCTGA
  51 CTTAGAAGGA GGTTCAGGCT ACGGTGAGCC GAAGCCACAC AGGAGCCATG
 101 GAAGTGGCAG AGCCCAGCAG CCCCACTGAA GAGGAGGAGG AGGAAGAGGA
 151 GCACTCGGCA GAGCCTCGGC CCCGCACTCG CTCCAATCCT GAAGGGGCTG
 201 AGGACCGGGC AGTAGGGGCA CAGGCCAGCG TGGGCAGCCG CAGCGAGGGT
 251 GAGGGTGAGG CCGCCAGTGC TGATGATGGG AGCCTCAACA CTTCAGGAGC
 301 CGGCCCTAAG TCCTGGCAGG TGCCCCCGCC AGCCCCTGAG GTCCAAATTC
 351 GGACACCAAG GGTCAACTGT CCAGAGAAAG TGATTATCTG CCTGGACCTG
 401 TCAGAGGAAA TGTCACTGCC AAAGCTGGAG TCGTTCAACG GCTCCAAAAC
 451 CAACGCCCTC AATGTCTCTC AGAAGATGAT TGAGATGTTC GTGCGGACAA
 501 AACACAAGAT CGACAAAAGC CACGAGTTTG CACTGGTGGT GGTGAACGAT
 551 GACACGGCCT GGCTGTCTGG CCTGACCTCC GACCCCCGCG AGCTCTGTAG
 601 CTGCCTCTAT GATCTGGAGA CGGCCTCCTG TTCCACCTTC AATCTGGAAG
 651 GACTTTTCAG CCTCATCCAG CAGAAAACTG AGCTTCCGGT CACAGAGAAC
 701 GTGCAGACGA TTCCCCCGCC ATATGTGGTC CGCACCATCC TTGTCTACAG
 751 CCGTCCACCT TGCCAGCCCC AGTTCTCCTT GACGGAGCCC ATGAAGAAAA
 801 TGTTCCAGTG CCCATATTTC TTCTTTGACG TTGTTTACAT CCACAATGGC
 851 ACTGAGGAGA AGGAGGAGGA GATGAGTTGG AAGGATATGT TTGCCTTCAT
 901 GGGCAGCCTG GATACCAAGG GTACCAGCTA CAAGTATGAG GTGGCACTGG
 951 CTGGGCCAGC CCTGGAGTTG CACAACTGCA TGGCGAAACT GTTGGCCCAC
1001 CCCCTGCAGC GGCCTTGCCA GAGCCATGCT TCCTACAGCC TGCTGGAGGA
1051 GGAGGATGAA GCCATTGAGG TTGAGGCCAC TGTCTGAACC ATCCCTGTAC
1101 ATCTGCACCT TCTTGTGCAA GGAAGTCCTT GGCCTAAAGC CTTGGTTCTC
1151 AAACTGGGTT CCTTGGGACC TCCGGGGTGG GGGGGTTCCA GGAGGCACGT
1201 AGGGTACCTT GCAGGGTCCT AGGAGGGAAA CCCAGGATTC CAGGAGGGAT
1251 CCCAGGAACT GTGGGCACCC ATTTTCTGTG TCTCCCAGCC CATTTCCACT
1301 CCTAGTTTGT CATGGATAAT TTTTGTTCTT CCCTGTGTGA TTTTTGCCAT
1351 CAAAATAAAA ATTTGAGACT CGTTAAAAAA AAAAAAAAAA AAAAAAAAAA
1401 AAAAAAAAAA AAAAAAAAAA AAAAAAGAAA AAAAAAAAAA AAAAAAA
```

BLAST Results

Entry HS806352 from database EMBL:
human STS EST192543.
Score = 1285, P = 2.5e-51, identities = 263/266

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 98 bp to 1084 bp; peptide length: 329
Category: similarity to unknown protein
Classification: no clue

```
   1 MEVAEPSSPT EEEEEEEEHS AEPRPRTRSN PEGAEDRAVG AQASVGSRSE
```

```
 51 GEGEAASADD GSLNTSGAGP KSWQVPPPAP EVQIRTPRVN CPEKVIICLD
101 LSEEMSLPKL ESFNGSKTNA LNVSQKMIEM FVRTKHKIDK SHEFALVVVN
151 DDTAWLSGLT SDPRELCSCL YDLETASCST FNLEGLFSLI QQKTELPVTE
201 NVQTIPPPYV VRTILVYSRP PCQPQFSLTE PMKKMFQCPY FFFDVVYIHN
251 GTEEKEEEMS WKDMFAFMGS LDTKGTSYKY EVALAGPALE LHNCMAKLLA
301 HPLQRPCQSH ASYSLLEEED EAIEVEATV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78n23, frame 2

PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana, N = 1, Score = 142, P = 1.5e-07

```
>PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana
          Length = 264

  HSPs:

 Score = 142 (21.3 bits), Expect = 1.5e-07, P = 1.5e-07
 Identities = 56/216 (25%), Positives = 97/216 (44%)

Query:    93 EKVIICLDL-SEEMSLPKLESFNGSKTNALNVSQKMIEMFVRTKHKIDKSHEFALVVVND 151
             E ++IC+D+ +E M   K    NG     +   ++ I +F+  K  I+  H FA   +
Sbjct:    26 EDILICIDVDAESMVEMKTTGTNGRPLIRMECVKQAIILFIHNKLSINPDHRFAFATLAK 85

Query:   152 DTAWLSG-LTSDPRELCSCLYDLE-TASCSTFNLEGLFSLIQQKTELPVTENVQTIPPPY 209
               AWL    TSD    + L  L    S S  +L  LF    Q+ ++    +N
Sbjct:    86 SAAWLKKEFTSDAESAVASLRGLSGNKSSSRADLTLLFRAAAQEAKVSRAQN-------R 138

Query:   210 VVRTILVYSRPPCQPQFSLTEPMKKMFQCPYFFFDVVYIHNGTEEKEEEMSWKDMF-AFM 268
             + R IL+Y R   +P       P+ +       F  DV+Y+H    ++     +D++ + +
Sbjct:   139 IFRVILIYCRSSMRPTHEW--PLNQKL----FTLDVMYLH---DKPSPDNCPQDVYDSLV 189

Query:   269 GSLD--TKGTSYKYEVALAGPALELHNCMAKLLAHPLQRPCQ 308
              +++   ++    Y +E     G A  +   M+ LL HP QR  Q
Sbjct:   190 DAVEHVSEYEGYIFESG-QGLARSVFKPMSMLLTHPQQRCAQ 230
```

Pedant information for DKFZphfbr2_78n23, frame 2

Report for DKFZphfbr2_78n23.2

```
[LENGTH]    329
[MW]        36560.10
[pI]        4.60
[HOMOL]     PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana 7e-07
[KW]        Alpha_Beta
[KW]        LOW_COMPLEXITY    9.73 %

SEQ  MEVAEPSSPTEEEEEEEEHSAEPRPRTRSNPEGAEDRAVGAQASVGSRSEGEGEAASADD
SEG  .xxxxxxxxxxxxxxxxxxxxx......................................
PRD  ccccccccchhhhhhhhhhcccccccccccchhhhhhhhhhcccccccccccccc

SEQ  GSLNTSGAGPKSWQVPPPAPEVQIRTPRVNCPEKVIICLDLSEEMSLPKLESFNGSKTNA
SEG  ............................................................
PRD  cccccccccccccccccccceeeccccccccceeeeeccccccccccccccccccccee

SEQ  LNVSQKMIEMFVRTKHKIDKSHEFALVVVNDDTAWLSGLTSDPRELCSCLYDLETASCST
SEG  ............................................................
PRD  ehhhhhhhhhhhhhhcccccceeeeeecchhhhhcccccchhhhhhhhhcccccccc

SEQ  FNLEGLFSLIQQKTELPVTENVQTIPPPYVVRTILVYSRPPCQPQFSLTEPMKKMFQCPY
SEG  ............................................................
PRD  hhhhhhhhhhhhhhhhhhhhhcccccccccceeeeeeeccccccccccccchhhhhheeee

SEQ  FFFDVVYIHNGTEEKEEEMSWKDMFAFMGSLDTKGTSYKYEVALAGPALELHNCMAKLLA
SEG  ............................................................
PRD  eeeeeeeecccchhhhhhhhhhhhhhhhhccccccccceeeeecccccchhhhhhhhhhh

SEQ  HPLQRPCQSHASYSLLEEEDEAIEVEATV
SEG  ................xxxxxxxxxx...
PRD  hccccccccchhhhhhhhhhhhhhhccc
```

(No Prosite data available for DKFZphfbr2_78n23.2)

(No Pfam data available for DKFZphfbr2_78n23.2)

DKFZphfbr2_7a24
----

group: brain derived

DKFZphfbr2_7a24 encodes a novel 142 amino acid protein with similarity to the C-terminal part of transforming growth factor-beta activated kinases.

The novel protein shows only similarity to the C-terminus of such kinases; no kinase domain is present.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to C-terminus of TGF-beta-activated kinase

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1697 bp
No poly A stretch found, no polyadenylation signal found

```
   1 GGGGAGAGAG GGGTTGTGAA GGGAAGCGGA AGGGAAGGGA AGGGAGGTCC
  51 CGTGGGACGC TGGGGTCTGG GGTAGAGCAG GTAGCAGCGT GCTGCCCTGA
 101 CAGCTGTCTC CGCTCCTCAG ATTGTCAGTG GCTGCTATGC AGCAGGTGCA
 151 GCCTGGTCTC TCACTGAGTC TCTACTCCAC AAAGGCAACG ACTGGCCAAG
 201 GCAGTGGCTG GCTCTGGGTT ACACAAGTGC AGACACTCAA CTAAGTGAGC
 251 TGGAAGACCC AGGAGAAGGC GGAGGCTCAG GTGCCCACAT GATCAGCACA
 301 GCCAGGGTAC CTGCTGACAA GCCTGTACGC ATCGCCTTTA GCCTCAATGA
 351 CGCCTCAGAT GATACACCCC CTGAAGACTC CATTCCTTTG GTCTTTCCAG
 401 AATTAGACCA GCAGCTACAG CCCCTGCCGC CTTGTCATGA CTCCGAGGAA
 451 TCCATGGAGG TGTTCAGACA GCACTGCCAA ATAGCAGAAG AATACCTTGA
 501 GGTCAAAAAG GAAATCACCC TGCTTGAGCA AAGGAAGAAG GAGCTCATTG
 551 CCAAGTTAGA TCAGGCAGAA GAGGAGAAGG TGGATGCTGC TGAGCTGGTT
 601 CGGGAATTCG AGGCTCTGAC GGAGGAGAAT CGGACGTTGA GGTTGGCCCA
 651 GTCTCAATGT GTGGAACAAC TGGAGAAACT TCGAATACAG TATCAGAAGA
 701 GGCAGGGCTC GTCCTAACTT TAAATTTTTC AGTGTGAGCA TACGAGGCTG
 751 ATGACTGCCC TGTGCTGGCC AAAAGATTTT TATTTTAAAT GAATAGTGAG
 801 TCAGATCTAT TGCTTCTCTG TATTACCCAC ATGACAACTG TCTATAATGA
 851 GTTTACTGCT TGCCAGCTTC TAGCTTGAGA GAAGGGATAT TTTAAATGAG
 901 ATCATTAACG TGAAACTATT ACTAGTATAT GTTTTTGGAG ATCAGAATTC
 951 TTTTCCAAAG ATATATGTTT TTTTCTTTTT TAGGAAGATA TGATCATGCT
1001 GTACAACAGG GTAGAAAATG GTAAAAATAG ACTATTGACT GACCCAGCTA
1051 AGAATCGCGG GCTGAGCAGA GTTAAACCAT GGGACAAACC CATAACATGT
1101 TCACCATAGT TTCACGTATG TGTATTTTTA AATTTCATGC CTTTAATATT
1151 TCAAATATGC TCAAATTTAA ACTGTCAGAA ACTTCTCTGC ATGTATTTAT
1201 ATTTGCCAGA GTATAAACTT TTATACTCTG ATTTTTATCC TTCAATGATT
1251 GATTATACTA AGAATAAATG GTCACATATC CTAAAAGCTT CTTCATGAAA
1301 TTATTAGCAG AAACCATGTT TGAAACCAAA GCACATTTGC CAATGCTAAC
1351 TGGCTGTTGT AATAATAAAC AGATAAGGCT GCATTTGCTT CATGCCATGT
1401 GACCTCACAG TAAACATCTC TGCCTTTGCC TGTGTGTGTT CTGGGGGAGG
1451 GGGGACATGG AAAAATATTG TTTGGACATT ACTTGGGTGA GTGCCCATGA
1501 AGACATCAGT GAACTTGTAA CTATTGTTTT GTTTTGGATT TAAGGAGATG
1551 TTTTAGATCA GTAACAGCTA ATAGGAATAT GCGAGTAAAT TCAGAATTGA
1601 AACAATTTCT CCTTGTTCTA CCTATCACCA CATTTTCTCA AATTGAACTC
1651 TTTGTTATAT GTCCATTTCT ATTCATGTAA CTTCTTTTTC ATTAAAC
```

BLAST Results

No BLAST result

Medline entries

98130593:
Role of TAK1 and TAB1 in BMP signaling in early Xenopus development.

Peptide information for frame 1

```
ORF from 289 bp to 714 bp; peptide length: 142
Category: similarity to known protein

    1 MISTARVPAD KPVRIAFSLN DASDDTPPED SIPLVFPELD QQLQPLPPCH
   51 DSEESMEVFR QHCQIAEEYL EVKKEITLLE QRKKELIAKL DQAEEEKVDA
  101 AELVREFEAL TEENRTLRLA QSQCVEQLEK LRIQYQKRQG SS
```

BLASTP hits

```
Entry U92030_1 from database TREMBL:
product: "TAK1"; Xenopus laevis TGF-beta-activated kinase TAK1 mRNA,
complete cds.
Score = 343, P = 1.3e-30, identities = 69/143, positives = 104/143

Entry AB009356_1 from database TREMBL:
product: "TGF-beta activated kinase 1a"; Homo sapiens mRNA for
TGF-beta activated kinase 1a, complete cds.
Score = 339, P = 2.6e-30, identities = 67/143, positives = 104/143

Entry MMPK_1 from database TREMBL:
product: "TAK1 (TGF-beta-activated kinase)"; Mouse mRNA for TAK1
(TGF-beta-activated kinase), complete cds.
Score = 339, P = 2.6e-30, identities = 67/143, positives = 104/143

Entry AB009357_1 from database TREMBL:
product: "TGF-beta activated kinase 1b"; Homo sapiens mRNA for
TGF-beta activated kinase 1b, complete cds.
Score = 339, P = 3.2e-30, identities = 67/143, positives = 104/143

Entry AB009358_1 from database TREMBL:
product: "TGF-beta activated kinase 1c"; Homo sapiens mRNA for
TGF-beta activated kinase 1c, complete cds.
Score = 144, P = 3.8e-09, identities = 30/67, positives = 47/67
```

Alert BLASTP hits for DKFZphfbr2_7a24, frame 1

```
PIR:JC5955 transforming growth factor-beta activated kinase (EC
-.-.-.-) 1a - Human, N = 1, Score = 339, P = 3e-30


>PIR:JC5955 transforming growth factor-beta activated kinase (EC -.-.-.-) 1a
     - Human
          Length = 579

  HSPs:

 Score = 339 (50.9 bits), Expect = 3.0e-30, P = 3.0e-30
 Identities = 67/143 (46%), Positives = 104/143 (72%)

Query:      1 MISTARVPADKPVRI-AFSLNDASDDTPPEDSIPLVFPELDQQLQPLPPCHDSEESMEVF 59
              MI+T+    ++KP R   ++  +D++D    ++SIP+  +   LD QLQPL PC +S+ESM VF
Sbjct:    437 MITTSGPTSEKPTRSHPWTPDDSTDTNGSDNSIPMAYLTLDHQLQPLAPCPNSKESMAVF 496

Query:     60 RQHCQIAEEYLEVKKEITLLEQRKKELIAKLDQAEEEKVDAAELVREFEALTEENRTLRL 119
              QHC++A+EY++V+ EI LL QRK+EL+A+LDQ E+++ + + LV+E + L +EN++L
Sbjct:    497 EQHCKMAQEYMKVQTEIALLLQRKQELVAELDQDEKDQQNTSRLVQEHKKLLDENKSLST 556

Query:    120 AQSQCVEQLEKLRIQYQKRQGSS 142
                QC +QLE +R Q QKRQG+S
Sbjct:    557 YYQQCKKQLEVIRSQQQKRQGTS 579
```

Pedant information for DKFZphfbr2_7a24, frame 1

Report for DKFZphfbr2_7a24.1

```
[LENGTH]        142
[MW]            16377.53
[pI]            4.64
[HOMOL]         TREMBL:U92030_1 product: "TAK1"; Xenopus laevis TGF-beta-activated kinase TAK1
mRNA, complete cds. 6e-26
[PROSITE]       CK2_PHOSPHO_SITE       3
```

```
[PROSITE]       PKC_PHOSPHO_SITE        2
[PROSITE]       ASN_GLYCOSYLATION       1
[PFAM]          TNFR/NGFR cysteine-rich region
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY      7.04 %
[KW]            COILED_COIL        33.10 %

SEQ     MISTARVPADKPVRIAFSLNDASDDTPPEDSIPLVFPELDQQLQPLPPCHDSEESMEVFR
SEG     ....................................xxxxxxxxxx............
PRD     ccccccccccccccccccccccccccccccccccchhhhhhhhcccccccccchhhhhhh
COILS   ............................................................

SEQ     QHCQIAEEYLEVKKEITLLEQRKKELIAKLDQAEEEKVDAAELVREFEALTEENRTLRLA
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhh
COILS   ...CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC..........

SEQ     QSQCVEQLEKLRIQYQKRQGSS
SEG     ......................
PRD     hhhhhhhhhhhhhhhhhhhccc
COILS   ......................
```

Prosite for DKFZphfbr2_7a24.1

PS00001	114->118	ASN_GLYCOSYLATION	PDOC00001
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00006	18->22	CK2_PHOSPHO_SITE	PDOC00006
PS00006	26->30	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006

Pfam for DKFZphfbr2_7a24.1

```
HMM_NAME     TNFR/NGFR cysteine-rich region

HMM             *CpeGtYtDWNHvpqClpCtrCePEMGQYMvqPCTwTQNTVC*
                 C++++ + +    +Q   C++ E+   ++++++ T   + ++
Query        49  CHDSEESMEVF-RQH--CQIAEE--YLEVKKEITLLEQRKK      84
```

DKFZphfbr2_7e22
----

group: brain derived

DKFZphfbr2_7e22.2 encodes a novel 286 amino acid protein similar to b561 cytochromes

The new protein shows strong similarity to B561 cytochromes, but contains no heme binding site. In addition, a myc-type, helix-loop-helix dimerization domain domain is present. This helix-loop-helix domain mediates protein dimerization and has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to cytochrome b561

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 4254 bp
Poly A stretch at pos. 4234, polyadenylation signal at pos. 4217

```
   1 GGGGACTACC CAGAGGGCTG CCGCCGCCTC TCCAAGTTCT TGTGGCCCCC
  51 GCGGTGCGGA GTATGGGGCG CTGATGGCCA TGGAGGGCTA CCGGCGCTTC
 101 CTGGCGCTGC TGGGGTCGGC ACTGCTCGTC GGCTTCCTGT CGGTGATCTT
 151 CGCCCTCGTC TGGGTCCTCC ACTACCGAGA GGGGCTTGGC TGGGATGGGA
 201 GCGCACTAGA GTTTAACTGG CACCCAGTGC TCATGGTCAC CGGCTTCGTC
 251 TTCATCCAGG GCATCGCCAT CATCGTCTAC AGACTGCCGT GGACCTGGAA
 301 ATGCAGCAAG CTCCTGATGA AATCCATCCA TGCAGGGTTA AATGCAGTTG
 351 CTGCCATTCT TGCAATTATC TCTGTGGTGG CCGTGTTTGA GAACCACAAT
 401 GTTAACAATA TAGCCAATAT GTACAGTCTG CACAGCTGGG TTGGACTGAT
 451 AGCTGTCATA TGCTATTTGT TACAGCTTCT TTCAGGTTTT TCAGTCTTTC
 501 TGCTTCCATG GGCTCCGCTT TCTCTCCGAG CATTTCTCAT GCCCATACAT
 551 GTTTATTCTG GAATTGTCAT CTTTGGAACA GTGATTGCAA CAGCACTTAT
 601 GGGATTGACA GAGAAACTGA TTTTTTCCCT GAGAGATCCT GCATACAGTA
 651 CATTCCCGCC AGAAGGTGTT TTCGTAAATA CGCTTGGCCT TCTGATCCTG
 701 GTGTTCGGGG CCCTCATTTT TTGGATAGTC ACCAGACCGC AATGGAAACG
 751 TCCTAAGGAG CCAAATTCTA CCATTCTTCA TCCAAATGGA GGCACTGAAC
 801 AGGGAGCAAG AGGTTCCATG CCAGCCTACT CTGGCAACAA CATGGACAAA
 851 TCAGATTCAG AGTTAAACAA TGAAGTAGCA GCAAGGAAAA GAAACTTAGC
 901 TCTGGATGAG GCTGGGCAGA GATCTACCAT GTAAAATGTT GTAGAGATAG
 951 AGCCATATAA CGTCACGTTT CAAAACTAGC TCTACAGTTT TGCTTCTCCT
1001 ATTAGCCATA TGATAATTGG GCTATGTAGT ATCAATATTT ACTTTAATCA
1051 CAAAGGATGG TTTCTTGAAA TAATTTGTAT TGATTGAGGC CTATGAACTG
1101 ACCTGAATTG GAAAGGATGT GATTAATATA AATAATAGCA GATATAAATT
1151 GTGGTTATGT TACCTTTATC TTGTTGAGGA CCACAACATT AGCACGGTGC
1201 CTTGTGCAGA ATAGATACTC AATATGTGAA TATGTGTCTA CTAGTAGTTA
1251 ATTGGATAAA CTGGCAGCAT CCCTGGCCTG TTGTCATGCA GTCATTTCCT
1301 GTTAATTCTG GGAGACAATG ATTTCACAAC TAGAGGGAAG CAGTCCTAAA
1351 AGTTTAAAAT CCGATAAGGA ATATCTGGGA CAGGGTTTAG ATCATGACTC
1401 TACACAGATA CCATGATGAG AGTATATTAA AGAAATTTAG GAAAGCACCT
1451 GGTTCCTTTC TCCCCATGCC TGCCTTCTGC TCCCTCCCCA GCTGGTTTGG
1501 GCTCAAATTG TCCCTGGAGA CTAGGGTTTA TGTTAGGGTA TTGATAGATT
1551 AGAGCAGGTG GTTGAAGAGA TCTTCTCTGG TCAGACTTGG AAGAATTTCC
1601 AAAAGTGAAG TTAGCCCCAA GACTTCCCTA GGGTTGATGT ACTTTATGAT
1651 CCAGATGCTA AACTTCTTAG AATGAAAATA TGCTTCAACA CTTAAGTAGC
1701 ATACACTGCC CTACAAACCT CAGAGAGCAC TTTTCCCCAA GTTCTTGTTT
1751 TTATTTTTGA AAGTACTCAC ACAGCACTTA CTATGCTCCA AACACTCCTC
1801 TAAGCACTTT ACACATATTA GCTCATTCAG TCCCCAGACA GACGGGATGA
1851 AGTAGGTATT GTTACTGTTC CCATTTTACA GGTGAGAGAT TTGAAGCCTG
1901 GGGAGGCTAG TAACTCACCC CAAGGTCACA CGGCTCATAC ATGGTGGGAC
1951 TGAGACTCAG ATGCAGGCAG TCTGGCACCT CAGTCTGGAT TCTAACCATT
2001 TCACTAAGCT ATTTTTGTCT TGTACTACTT TGACCCACCC CTGAATAAAC
2051 CTCAATTGCT GGAGTGGGGT GTAGTTATTA AAGGGATGCT TTTTACCTTT
2101 TGCTGTCTGC TGTGGCAGAT TCCCCAGATA ACCAAGGAAA AGGGGCCACC
2151 CATACCTGGA AATAGGCCAT AGGGCCCCTA CTACTGCCAA CAAGCCATGG
2201 CCTACCTTGA CACTTGTTTG ATCTTAAAAT TGTGTCTTGG TAACAAAAGA
2251 TTTGGACAGG CATATCTGTA GCTTTCAAGT TAATTAATTG CAATATTTTT
2301 TTCTTCAGGA TTTTAGCTGC TGAACAACTT TCAGTTTGGA GCTAAAAGAG
2351 ACCTGTCTCA TGGTCTGCCC TTCCCTGGGG CAATAGCTAG GGTCTTTCCT
2401 GATTTTTATG GAATTTTAGG GGATATTTTG AGCTTTGGGT TCTCAGTAGT
```

```
2451 GAATTGAGAC TTGGAGGTGA CTTTTCATGT TTGGAGTATC ATCTCTGTCT
2501 GGGCTCTGGG CTGACAAATT AAAACCTAGA GTAGTGCTTA TGCTGAAATG
2551 ATACTTTTCA TTTTTTGGTT GATTTTTTTG CCTTCCCTTC AATTTTAAAC
2601 TGAAGCATTT TAATGTGGGT AGAAACTCTA CACCAAATAC ACTAAACATT
2651 TTGGTGCTTA GTGGATTTCT TTTTAGGTAA CTGGTACTTA CTTCCAAAGA
2701 CTGAATACAA GCCACACTCC ATCATATCCC TTAAACTTCA TGAAAAACCA
2751 TTCAAGATCC CCTTGCTGCA ACACTGTTCT CTTCTTCTCT ACTAAATTCT
2801 ATTTCCAAAA TTGGTAATAG AGCCAGAAGG ATCCCCAGTA CCCAGCCCTC
2851 TGCCTGGCAC AAAGTGGTAG CACAATTAAA TTCAGTATGG GTGGAGCATG
2901 GTACAGTCTT GGTGCCATAG AAGGAGTAGT TGCATAGTCA CACATCATTT
2951 GATAAGTTGG ATGTTCCATT ACATAGAGGA ACACAAAATT CCAGGGTTTT
3001 TGGAGGAAGG GATTAGATAG CGACTAAGCC GCCAGAATTG AGGTGGCCAT
3051 TCCTTTTTGT ATAGGCTAAG AAACAGGTTA TCAGTGAAAA GTTAATTATG
3101 GCTTTGGCAC TAGAATAGCA CTGTTGCAAA GTATTTAAGC ACCCCCCATC
3151 TCAGCCCTTT ATTTTATCTT TCATGTGGGC TAATGTGAGG ATAATCTTAC
3201 AGATATTATA GGAATTTCTT TTCTATCTTT ATGAAAACAA CGTATATAAA
3251 ATATATCTAG AAAACCTTTG TTTGAGACTC TTATTTAATG GGCTTTTGAT
3301 TCTAATGATA ATTGTACCTT TATCTTTCAA AAGCTGATAT TTCCTACCTA
3351 AGCATCTCCC GAGAAAAATA TCTCATTAAA AAGCCCATAA ATAATAGGGG
3401 AGAAGAAAGC CTTAGGTATC AATTCCAAAA CAGTGATTGA AATTTCCCAA
3451 AATAATTATG GCTTCTGTCA TCTCCAGAGA TAATCTGGCT TGGTTTACCC
3501 CATAATCTAA TTTCAGAAAA GAAAGCTTTA TTTTAACACT CATCTGAATC
3551 AACATTAAAG CCTTTTCTCT CAAAGCGTTT ATTGAGAAAC TCAAATGAAT
3601 ATACTTTTTG AATTACTGTC ATCAAAAGTG TACGGCTTCC TGTGCTGCTT
3651 GTGTCAAATG GAACCTGCCC TCTAAAGCAC TTTCTTTCCT TTACTTGCGT
3701 GGTTTCATGT AAGCTGTGCT GTTTAGAAAC AACATCTCAG ACTTTACAAA
3751 GAAATGACAA AGAAGGCAAT TGCACTTTTT AAGGGATATC GACAAGCAGT
3801 TTCTGTTTTC TAAAGGACAA AATACAGAGT GTGTGTCATT TTTAATTAGA
3851 TTCTTTCCCC TGCTGAGTTG GAAATTCCAG TGCAGCACTG ATTGACCACA
3901 GTTGCCAATC TAAAAGCACA AAGACAGAAG TAAAGCTTTA TGCTAATTTT
3951 ATTTCAATAT GATAGAAAAT TTATCTTGGT ATGTCCTTTT TTAGATAACT
4001 CCAGCAGGAA ACTGTAACTG CTATGTCTTT AGGAAAACGT AGAAGAAAGA
4051 ACATTATTAT TCTTTAATTC CTACAAGGTA CTTGAAAACC TTAAGTGAAA
4101 AAGATTTCTA TCTTTTTATC TTGGCGCATT TATGGAAAAA ATATTAACTG
4151 TCCTGAATAT TTTATAATTT TGTAGGAAAA ATATGCATCT ATTTTTTCTT
4201 GACTTCTTTT ATATAGTAAT AAAAGTTATT TTGGAAAAAA AAAAAAAAAA
4251 AAAA
```

BLAST Results

Entry HSG20626 from database EMBL:
human STS A005Z27.
Score = 860, P = 3.0e-32, identities = 176/181

Medline entries

89030633:
The structure of cytochrome b561, a secretory vesicle-specific electron transport protein.

Peptide information for frame 2

ORF from 74 bp to 931 bp; peptide length: 286
Category: strong similarity to known protein
Classification: unset

```
  1 MAMEGYRRFL ALLGSALLVG FLSVIFALVW VLHYREGLGW DGSALEFNWH
 51 PVLMVTGFVF IQGIAIIVYR LPWTWKCSKL LMKSIHAGLN AVAAILAIIS
101 VVAVFENHNV NNIANMYSLH SWVGLIAVIC YLLQLLSGFS VFLLPWAPLS
151 LRAFLMPIHV YSGIVIFGTV IATALMGLTE KLIFSLRDPA YSTFPPEGVF
201 VNTLGLLILV FGALIFWIVT RPQWKRPKEP NSTILHPNGG TEQGARGSMP
251 AYSGNNMDKS DSELNNEVAA RKRNLALDEA GQRSTM
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_7e22, frame 2

SWISSPROT:C561_SHEEP CYTOCHROME B561 (CYTOCHROME B-561)., N = 1, Score

= 460, P = 1.3e-43

PIR:S01167 cytochrome b561 - bovine, N = 1, Score = 457, P = 2.7e-43

SWISSPROT:C561_PIG CYTOCHROME B561 (CYTOCHROME B-561)., N = 1, Score = 452, P = 9.1e-43

PIR:S53321 cytochrome B561 - human, N = 1, Score = 451, P = 1.2e-42

>SWISSPROT:C561_SHEEP CYTOCHROME B561 (CYTOCHROME B-561).
Length = 252

HSPs:

Score = 460 (69.0 bits), Expect = 1.3e-43, P = 1.3e-43
Identities = 96/218 (44%), Positives = 131/218 (60%)

```
Query:    18 LVGFLSVIFALVWVLHYREGLGWDGSALEFNWHPVLMVTGFVFIQGIAIIVYRLPWTWKC 77
             L+G   V     W+  YR G+ W+ SAL+FN HP+ MV G VF+QG A++VYR+
Sbjct:    23 LLGLTVVAMTGAWLGMYRGGIAWE-SALQFNVHPLCMVIGLVFLQGDALLVYRV--FRNE 79

Query:    78 SKLLMKSIHAGLNAVAAILAIISVVAVFENHNVNNIANMYSLHSWVGLIAVICYLLQLLS 137
             +K    K +H  L+  A ++A++ +VAVFE+H        A++YSLHSW G++     + Q L
Sbjct:    80 AKRTTKVLHGLLHVFAFVIALVGLVAVFEHHRKKGYADLYSLHSWCGILVFALFFAQWLV 139

Query:   138 GFSVFLLPWAPLSLRAFLMPIHVYSGIVIFGTVIATALMGLTEKLIFSLRDPAYSTFPPE 197
             GFS FL P A  SLR+   P HV+ G  IF   +ATAL+GL E L+F L     YSTF PE
Sbjct:   140 GFSFFLFPGASFSLRSRYRPQHVFFGAAIFLLSVATALLGLKEALLFEL-GTKYSTFEPE 198

Query:   198 GVFVNTLGLLILVFGALIFWIVTRPQWKRPKEPNSTIL 235
             GV  N LGLL+  F  ++ +I+TR  WKRP +      L
Sbjct:   199 GVLANVLGLLLAAFATVVLYILTRADWKRPLQAEEQAL 236
```

Pedant information for DKFZphfbr2_7e22, frame 2

Report for DKFZphfbr2_7e22.2

```
[LENGTH]    286
[MW]        31638.58
[pI]        9.12
[HOMOL]     SWISSPROT:C561_SHEEP CYTOCHROME B561 (CYTOCHROME B-561). 4e-40
[PIRKW]     transmembrane protein 9e-40
[KW]        SIGNAL_PEPTIDE 40
[KW]        TRANSMEMBRANE  5
[KW]        LOW_COMPLEXITY      4.90 %
```

```
SEQ   MAMEGYRRFLALLGSALLVGFLSVIFALVWVLHYREGLGWDGSALEFNWHPVLMVTGFVF
SEG   ............................................................
PRD   ccchhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhcccccccccccccccchhhhhhhhh
MEM   ................................................MMMMMMMMMMMM

SEQ   IQGIAIIVYRLPWTWKCSKLLMKSIHAGLNAVAAILAIISVVAVFENHNVNNIANMYSLH
SEG   ...........................xxxxxxxxxxxxxx...................
PRD   cccccceeeeeccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccceeecc
MEM   MMMMMMMMMMMMMM........MMMMMMMMMMMMMMMMMMMMMMMMM.............

SEQ   SWVGLIAVICYLLQLLSGFSVFLLPWAPLSLRAFLMPIHVYSGIVIFGTVIATALMGLTE
SEG   ............................................................
PRD   ccchhhhhhhhhhhhhhheeeeccccccccccccccceeeeeeeeeehhhhhhhhhh
MEM   ....MMMMMMMMMMMMMMMMMMMMM...........MMMMMMMMMMMMMMMMMMMMM...

SEQ   KLIFSLRDPAYSTFPPEGVFVNTLGLLILVFGALIFWIVTRPQWKRPKEPNSTILHPNGG
SEG   ............................................................
PRD   hhhhhhhccccccccccchhhhhhhhhhhhhhhhheeeeeeccccccccccccccccccc
MEM   ................MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...............

SEQ   TEQGARGSMPAYSGNNMDKSDSELNNEVAARKRNLALDEAGQRSTM
SEG   ..............................................
PRD   cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcccc
MEM   ..............................................
```

(No Prosite data available for DKFZphfbr2_7e22.2)

(No Pfam data available for DKFZphfbr2_7e22.2)

DKFZphfbr2_7j4
----

group: brain derived

DKFZphfbr2_7j4 encodes a novel 233 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, 1 EST hit

Sequenced by GBF

Locus: unknown

Insert length: 1050 bp
Poly A stretch at pos. 1027, polyadenylation signal at pos. 1007

```
   1 GGGGACACAA AGGGGTGGTC ACCCTGCCCT CACCTTGACC TGTAAGTTGC
  51 CTAGGACAGT GGCCTGGTCC CAGGGGCTGT TGTGGGGAGT TGAAGAACAC
 101 CCTGGCCTCC TCCATCATGT CGGCCAAGAG GGCAGAATTG AAGAAAAACAC
 151 ATCTGTGCAA GAACTACAAG GCAGTTTGCC TGGAATTGAA GCCAGAGCCG
 201 ACCAAAACAT TTGATTACAA AGCAGTTAAA CAAGAAGGGC GGTTTACCAA
 251 AGCAGGAGTG ACACAGGACC TAAAGAATGA ACTCAGGGAA GTGAGAGAAG
 301 AGCTCAAGGA GAAAATGGAG GAGATAAAAC AGATAAAGGA TCTAATGGAC
 351 AAGGATTTTG ATAAACTTCA CGAATTTGTG GAAATTATGA AGGAAATGCA
 401 GAAAGATATG GATGAGAAGA TGGACATTTT AATAAATACA CAGAAGAACT
 451 ATAAGCTTCC CCTTAGAAGA GCACCAAAGG AGCAGCAGGA ACTCAGGCTG
 501 ATGGGAAAGA CTCACAGAGA ACCACAGCTC AGGCCCAAGA AAATGGATGG
 551 AGCCAGTGGA GTCAATGGAG CACCCTGTGC TCTTCACAAG AAGACGATGG
 601 CACCACAAAA AACAAAACAG GGCTCACTGG ATCCCCTTCA TCACTGTGGG
 651 ACCTGCTGCG AGAAATGTTT GTTGTGTGCT CTAAAGAACA ACTACAATCG
 701 GGGGAACATT CCTTCAGAGG CCTCAGGCCT TTACAAAGGT GGAGAGGAGC
 751 CAGTGACCAC CCAACCTTCT GTGGGCCACG CTGTGCCTGC CCCAAAGTCC
 801 CAGACTGAGG GAAGGTGAAG CTTAACTGCC AGCTTGAAAT GAGAGTAAAG
 851 AAGATACAGA GCAAACAGTG TTTCAGAAAC TGTCCTGCCC TGGGTGTGAT
 901 TCTTTGGCTT CAATTTGAAG GAGGAGGAAT GATGGGATTT CATATTTTAT
 951 TTCACACCAG TTCCTCCTTG TTTCATCTCT TTGCTAAGCT GGCTGCTTCT
1001 ACCATCTAAT AAATAATTGG CCAAGTTAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 117 bp to 815 bp; peptide length: 233
Category: putative protein

```
  1 MSAKRAELKK THLCKNYKAV CLELKPEPTK TFDYKAVKQE GRFTKAGVTQ
 51 DLKNELREVR EELKEKMEEI KQIKDLMDKD FDKLHEFVEI MKEMQKDMDE
101 KMDILINTQK NYKLPLRRAP KEQQELRLMG KTHREPQLRP KKMDGASGVN
151 GAPCALHKKT MAPQKTKQGS LDPLHHCGTC CEKCLLCALK NNYNRGNIPS
201 EASGLYKGGE EPVTTQPSVG HAVPAPKSQT EGR
```

BLASTP hits

Entry JC2223 from database PIR:
major surface glycoprotein 3 - Pneumocystis carinii (fragment)
Score = 109, P = 3.5e-04, identities = 41/136, positives = 67/136

Alert BLASTP hits for DKFZphfbr2_7j4, frame 3

TREMBLNEW:PCP115C_1 product: "P115C"; Pneumocystis carinii mRNA for P115C, partial sequence., N = 1, Score = 109, P = 0.00024

```
>TREMBLNEW:PCP115C_1 product: "P115C"; Pneumocystis carinii mRNA for P115C,
     partial sequence.
            Length = 196

  HSPs:

 Score = 109 (16.4 bits), Expect = 2.4e-04, P = 2.4e-04
 Identities = 41/134 (30%), Positives = 67/134 (50%)

Query:    14 CKN-YKAVCLELKPEPTKTFDYKAVKQEGRFTKA-GVTQDLKNELREVREELKEKMEEIK 71
             CK   K  C ELK        + K VK+    TK  G  ++LK+++++  E  KE++E  K
Sbjct:    22 CKTELKKYCEELKEADGLKVNDK-VKEICDDTKRDGKCKELKDKVKKELETFKEELE--K 78

Query:    72 QIKDLMDKDFDKLHEFVEIMKEMQKDMDEKMDILINTQKNYKLPLRRAPKEQQELRLMGK 131
              +KD+ D++ +K  E    +++E   D D K + +   +  YKL  +R    E    LR +GK
Sbjct:    79 ALKDIKDENCEKYEEKCILLEETNHD-DVKKNCVKLREGCYKLKRKRVA-EDLLLRALGK 136

Query:   132 THREPQLRPKKMDGAS 147
              +  +    K  D  S
Sbjct:   137 DVKNGECEKKMKDVCS 152
```

Pedant information for DKFZphfbr2_7j4, frame 3

Report for DKFZphfbr2_7j4.3

```
[LENGTH]        233
[MW]            26533.95
[pI]            9.18
[PROSITE]       MYRISTYL        3
[PROSITE]       CK2_PHOSPHO_SITE        3
[PROSITE]       PKC_PHOSPHO_SITE        3
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY      14.59 %
[KW]            COILED_COIL         13.73 %

SEQ    MSAKRAELKKTHLCKNYKAVCLELKPEPTKTFDYKAVKQEGRFTKAGVTQDLKNELREVR
SEG    ...................................................xxxxxxxxx
PRD    ccchhhhhhhhhhcccchhhhhhhccccccccccccceeecccccccccccccchhhhhhhhh
COILS  ................................................cccccccccccc

SEQ    EELKEKMEEIKQIKDLMDKDFDKLHEFVEIMKEMQKDMDEKMDILINTQKNYKLPLRRAP
SEG    xxxxxxxxx...................xxxxxxxxxxxxxxx.................
PRD    hhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhchhhhhhhhhccccccccccc
COILS  CCCCCCCCCCCCCCCCCCC.........................................

SEQ    KEQQELRLMGKTHREPQLRPKKMDGASGVNGAPCALHKKTMAPQKTKQGSLDPLHHCGTC
SEG    ............................................................
PRD    hhhhhhhhhcccccccccccccccccccccccccchhhhhhcccccccccccccccccccc
COILS  ............................................................

SEQ    CEKCLLCALKNNYNRGNIPSEASGLYKGGEEPVTTQPSVGHAVPAPKSQTEGR
SEG    .....................................................
PRD    chhhhhhhccccccccccccccccccccccccccccccccccccccccccccc
COILS  .....................................................
```

Prosite for DKFZphfbr2_7j4.3

PS00005	2->5	PKC_PHOSPHO_SITE	PDOC00005
PS00005	108->111	PKC_PHOSPHO_SITE	PDOC00005
PS00005	132->135	PKC_PHOSPHO_SITE	PDOC00005
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	179->183	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00008	151->157	MYRISTYL	PDOC00008
PS00008	196->202	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_7j4.3)

DKFZphfbr2_82c20
----

group: transmembrane protein

DKFZphfbr2_82c20 encodes a novel 492 amino acid protein with very weak similarity to C. elegans cosmid D1007.

The novel protein contains 7 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans D1007.5 ;
membrane regions: 7
Summary DKFZphfbr2_82c20 encodes a novel 492 amino acid protein with similarity to a hypothetical C.elegans protein.

similarity to C.elegans D1007.5

complete cDNA (Bp 1-100 GC ritch), complete cds,
potential start at Bp 128 matches Kozak consensus PyNNatgG,
EST hits, localisation? primer B of STS doesn't match perfect!
TRANSMEMBRANE 7

Sequenced by DKFZ

Locus: /map="109.9 cR from top of Chr1 linkage group"???

Insert length: 1804 bp
Poly A stretch at pos. 1794, no polyadenylation signal found

```
   1 CGGCGGGAGC GCGCGGCTGA TACCCGGGAC TGGGCTGCGG CGGTTAGTCC
  51 TCTCCCGGCC GCCGTCGCCT CCGACATATT GCTCGCAGGA GCTGCGGCGG
 101 CGAAGCGGAG AGCACCGGGG GGAGGAGATG GGAGGACGAA GAGGTCCCAA
 151 CAGGACATCT TACTGTCGAA ATCCGCTCTG TGAGCCGGGA TCCTCGGGGG
 201 GCTCTAGTGG AAGCCACACT TCCAGTGCAT CGGTGACCAG TGTTCGTTCC
 251 CGCACCAGGA GCAGTTCTGG AACAGGCCTC TCCAGCCCTC CTCTGGCCAC
 301 CCAAACTGTT GTGCCTCTAC AGCACTGCAA GATCCCCGAG CTGCCAGTCC
 351 AGGCCAGCAT TCTGTTTGAG TTGCAGCTCT TCTTCTGCCA GCTCATAGCA
 401 CTCTTCGTCC ACTACATCAA CATCTACAAG ACAGTGTGGT GGTATCCACC
 451 TTCCCACCCA CCCTCCCACA CCTCCCTGAA CTTCCATCTG ATCGACTTCA
 501 ACTTGCTGAT GGTGACCACC ATCGTTCTGG GCCGCCGCTT CATTGGGTCC
 551 ATCGTGAAGG AGGCCTCTCA GAGGGGGAAG GTCTCCCTCT TTCGCTCCAT
 601 CCTGCTGTTC CTCACTCGCT TCACCGTTCT CACGGCAACA GGCTGGAGTC
 651 TGTGCCGATC CCTCATCCAC CTCTTCAGGA CCTACTCCTT CCTGAACCTC
 701 CTGTTCCTCT GCTATCCGTT TGGGATGTAC ATTCCGTTCC TGCAGCTGAA
 751 TTGCGACCTC CGCAAGACAA GCCTCTTCAA CCACATGGCC TCCATGGGGC
 801 CCCGGGAGGC GGTCAGTGGC CTGGCAAAGA GCCGGGACTA CCTCCTGACA
 851 CTGCGGGAGA CGTGGAAGCA GCACACAAGA CAGCTGTATG GCCCGGACGC
 901 CATGCCCACC CATGCCTGCT GCCTGTCACC CAGCCTCATC CGCAGTGAGG
 951 TGGAGTTCCT CAAGATGGAC TTCAACTGGC GCATGAAGGA AGTGCTCGTC
1001 AGCTCCATGC TGAGCGCCTA CTATGTGGCC TTTGTGCCTG TCTGGTTCGT
1051 GAAGAACACA CATTACTATG ACAAGCGCTG GTCCTGTGAA CTCTTCCTGC
1101 TGGTGTCCAT CAGCACCTCC GTGATCCTCA TGCAGCACCT GCTGCCTGCC
1151 AGCTACTGTG ACCTGCTGCA CAAGGCCGCC GCCCATCTGG GCTGTTGGCA
1201 GAAGGTGGAC CCAGCGCTGT GCTCCAACGT GCTGCAGCAC CCGTGGACTG
1251 AAGAATGCAT GTGGCCGCAG GGCGTGCTGG TGAAGCACAG CAAGAACGTC
1301 TACAAAGCCG TAGGCCACTA CAACGTGGCT ATCCCCTCTG ACGTCTCCCA
1351 CTTCCGCTTC CATTTCTTTT TCAGCAAACC TCTGCGGATC CTCAACATCC
1401 TCCTGCTGCT GGAGGGCGCT GTCATTGTCT ATCAGCTGTA CTCCCTAATG
1451 TCCTCTGAAA AGTGGCACCA GACCATCTCG CTGGCCCTCA TCCTCTTCAG
1501 CAACTACTAT GCCTTCTTCA AGCTGCTCCG GGACCGCTTG GTATTGGGCA
1551 AGGCCTACTC ATACTCTGCT AGCCCCCAGA GAGACCTGGA CCACCGTTTC
1601 TCCTGAGCCC TGGGGTCACC TCAGGGACAG CGTCCAGGCT TCAGCCAAGG
1651 GCTCCCTGGC AAGGGGCTGT TGGGTAGAAG TGGTGGTGGG GGGGACAAAA
1701 GACAAAAAAA TCCACCAGAG CTTTGTATTT TTGTTACGTA CTGTTTCTTT
1751 GATAATTGAT GTGATAAGGA AAAAAGTCCT ATTTTTATAC TCCCAAAAAA
1801 AAAA
```

BLAST Results

Entry HS285343 from database EMBL:
human STS WI-17488.

```
Score = 1225, P = 1.3e-50, identities = 263/281
```

Medline entries

No Medline entry

Peptide information for frame 2

```
  1 MGGRRGPNRT SYCRNPLCEP GSSGGSSGSH TSSASVTSVR SRTRSSSGTG
 51 LSSPPLATQT VVPLQHCKIP ELPVQASILF ELQLFFCQLI ALFVHYINIY
101 KTVWWYPPSH PPSHTSLNFH LIDFNLLMVT TIVLGRRFIG SIVKEASQRG
151 KVSLFRSILL FLTRFTVLTA TGWSLCRSLI HLFRTYSFLN LLFLCYPFGM
201 YIPFLQLNCD LRKTSLFNHM ASMGPREAVS GLAKSRDYLL TLRETWKQHT
251 RQLYGPDAMP THACCLSPSL IRSEVEFLKM DFNWRMKEVL VSSMLSAYYV
301 AFVPVWFVKN THYYDKRWSC ELFLLVSIST SVILMQHLLP ASYCDLLHKA
351 AAHLGCWQKV DPALCSNVLQ HPWTEECMWP QGVLVKHSKN VYKAVGHYNV
401 AIPSDVSHFR FHFFFSKPLR ILNILLLLEG AVIVYQLYSL MSSEKWHQTI
451 SLALILFSNY YAFFKLLRDR LVLGKAYSYS ASPQRDLDHR FS
```

ORF from 128 bp to 1603 bp; peptide length: 492
Category: similarity to unknown protein
Prosite motifs: LEUCINE_ZIPPER (210-232)
LEUCINE_ZIPPER (210-232)

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82c20, frame 2

TREMBL:CEAF3151_8 gene: "D1007.5"; Caenorhabditis elegans cosmid D1007., N = 2, Score = 247, P = 4.6e-29

```
>TREMBL:CEAF3151_8 gene: "D1007.5"; Caenorhabditis elegans cosmid D1007.
          Length = 512

 HSPs:

 Score = 247 (37.1 bits), Expect = 4.6e-29, Sum P(2) = 4.6e-29
 Identities = 58/204 (28%), Positives = 102/204 (50%)

Query:   291 VSSMLSAYYVAFVPVWFVKNTHYYDKRWSCELFLLVSISTSVILMQHLLPASYCDLLHKA 350
             +S ML   +V F       +         ++ W C+L ++V ++     + + +L P +Y DLLH+A
Sbjct:   299 LSIMLPCIFVPFKTSQGIPQKILINEVWECQLAIVVGLTAFSLYVAYLSPLNYLDLLHRA 358

Query:   351 AAHLGCWQKVD-PAL----CSNVLQHPWTEECMWPQGVLVKHSKN-VYKAVGHYNV---- 400
             A HLG W +++ P +        +      PW+E C++  G  V+       Y+A     ++
Sbjct:   359 AIHLGSWHQIEGPRIGHTGSMSSAPTPWSEFCLYNDGETVQMPDGRCYRAKSSNSIRTVA 418

Query:   401 AIPSDVSHFRFHFFFSKPLRILNILLLLEGAVIVYQLYSLMSSEKWHQTISLALILFSNY 460
             A P     H  F      KP  ++NI+     E  +I  Q + L+ +  +  W     ++  L++F+NY
Sbjct:   419 AHPESSRHNTFFKVLRKPNNLINIMCSFEFLLIFIQFWMLVLTNDWQHIVTFVLLMFANY 478

Query:   461 YAFFKLLRDRLVLGKAYSYSASPQRDL 487
                F KL +D+++L + Y   S   Q DL
Sbjct:   479 LLFAKLFKDKIILSRIYEPS---QEDL 502

 Score = 178 (26.7 bits), Expect = 4.3e-21, Sum P(2) = 4.3e-21
 Identities = 50/179 (27%), Positives = 90/179 (50%)

Query:   262 HACCLSPSLIRSEVEFLKMDFNWRMKEVLVSSMLSAYYVAFVPVWFV--KNTHYYDKR-- 317
             H C  SP+ IR E++ L   D     R+K+ + + + +A+     +P  FV   K +      ++
Sbjct:   262 HMCSDSPAQIREEIQVLIDDLVLRVKKSIFAGVSTAFLSIMLPCIFVPFKTSQGIPQKIL 321

Query:   318 ----WSCELFLLVSISTSVILMQHLLPASYCDLLHKAAAHLGCWQKVD-PAL----CSNV 368
                 W C+L ++V ++     + + +L P +Y DLLH+AA HLG W +++ P +        +
Sbjct:   322 INEVWECQLAIVVGLTAFSLYVAYLSPLNYLDLLHRAAIHLGSWHQIEGPRIGHTGSMSS 381

Query:   369 LQHPWTEECMWPQGVLVKHSKN-VYKAVGHYNV-AIPSDVSHFRFHFFFSKPLRILNILL 426
                PW+E C++  G  V+       Y+A     ++  + +     R + FF K LR  N L+
Sbjct:   382 APTPWSEFCLYNDGETVQMPDGRCYRAKSSNSIRTVAAHPESSRHNTFF-KVLRKPNNLI 440
```

```
 Score = 146 (21.9 bits), Expect = 4.6e-29, Sum P(2) = 4.6e-29
 Identities = 34/86 (39%), Positives = 50/86 (58%)

Query:     52 SSPPLATQTVVPLQHCKIPELP-VQASILFELQLFFCQLIALFVHYINIYKTVWWYPPSH 110
              +S P A+    + + H   P++    Q  + FE  LF    ++ALF+ Y+NIYKT+WW P S+
Sbjct:     19 ASIPRASGVTLSV-HPIWPDIQFTQGELFFECTLFLYSVLALFLQYLNIYKTLWWLPKSY 77

Query:    111 PPSHTSLNFHLIDFNLLMVTTIVLGRR 137
                  H SL FHLI+    L     ++LG R
Sbjct:     78 --WHYSLKFHLINPYFLSCVGLLLGWR 102

 Score = 39 (5.9 bits), Expect = 6.8e-18, Sum P(2) = 6.8e-18
 Identities = 12/41 (29%), Positives = 20/41 (48%)

Query:    154 LFRSILLFLTRFTVLTATGWSLCRSLIHLFRTYSFLNLLFL 194
              L+   + LFL ++   +   T W L +S  H     +  +N  FL
Sbjct:     53 LYSVLALFL-QYLNIYKTLWWLPKSYWHYSLKFHLINPYFL 92
```

Pedant information for DKFZphfbr2_82c20, frame 2

Report for DKFZphfbr2_82c20.2

```
[LENGTH]        492
[MW]            56274.05
[pI]            9.51
[HOMOL]         TREMBL:CEAF3151_8 gene: "D1007.5"; Caenorhabditis elegans cosmid D1007. 4e-31

[PROSITE]       LEUCINE_ZIPPER 1
[PROSITE]       AMIDATION      2
[PROSITE]       MYRISTYL       5
[PROSITE]       CAMP_PHOSPHO_SITE     2
[PROSITE]       CK2_PHOSPHO_SITE      3
[PROSITE]       GLYCOSAMINOGLYCAN     1
[PROSITE]       PKC_PHOSPHO_SITE      5
[PROSITE]       ASN_GLYCOSYLATION     1
[KW]            TRANSMEMBRANE  7
[KW]            LOW_COMPLEXITY     8.74 %

SEQ     MGGRRGPNRTSYCRNPLCEPGSSGGSSGSHTSSASVTSVRSRTRSSSGTGLSSPPLATQT
SEG     ....................xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.......
PRD     ccccccccccccccccccccccccccccccccccccccccccccceeccccccccccccee
MEM     ............................................................

SEQ     VVPLQHCKIPELPVQASILFELQLFFCQLIALFVHYINIYKTVWWYPPSHPPSHTSLNFH
SEG     ............................................................
PRD     eeecccccccccchhhhhhhhhhhhhhhhhhhhhhheeeeeccccccccccccccceeeeee
MEM     ............MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.........MMMMMMMMM

SEQ     LIDFNLLMVTTIVLGRRFIGSIVKEASQRGKVSLFRSILLFLTRFTVLTATGWSLCRSLI
SEG     ............................................................
PRD     eeehhhhhhhhhhhhheeeehhhhhhhcccchhhhhhhhhhhhhhhhhccccchhhhhhh
MEM     MMMMMMMMMMMMMMMMM.................MMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ     HLFRTYSFLNLLFLCYPFGMYIPFLQLNCDLRKTSLFNHMASMGPREAVSGLAKSRDYLL
SEG     ............................................................
PRD     hhhhhhhhheeeeeeecccccceeeeccccchhhhhhhhhhccchhhhhhhhhhhhhhhh
MEM     ............................................................

SEQ     TLRETWKQHTRQLYGPDAMPTHACCLSPSLIRSEVEFLKMDFNWRMKEVLVSSMLSAYYV
SEG     ............................................................
PRD     hhhhhhhhhhhcccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhh
MEM     ...........................................MMMMMMMMMMMMMMMMM

SEQ     AFVPVWFVKNTHYYDKRWSCELFLLVSISTSVILMQHLLPASYCDLLHKAAAHLGCWQKV
SEG     ............................................................
PRD     heeeeeeeecccccccchhhhhhhhhhhcchhhhhhhhhhcchhhhhhhhhhhhhhhhccc
MEM     MMMMMMMMM.........MMMMMMMMMMMMMMMMMMMMMMMMMMMMM.............

SEQ     DPALCSNVLQHPWTEECMWPQGVLVKHSKNVYKAVGHYNVAIPSDVSHFRFHFFFSKPLR
SEG     ..........................................................xx
PRD     ccccccccccccccceeeccccceeeeecccceeeeccccccccccccccceeeeeecccch
MEM     ..................................................MMMMMMMMMM

SEQ     ILNILLLLEGAVIVYQLYSLMSSEKWHQTISLALILFSNYYAFFKLLRDRLVLGKAYSYS
SEG     xxxxxxxx....................................................
PRD     hhhhhhhhhheeeeehhhhcccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
MEM     MMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMM..............
```

SEQ	ASPQRDLDHRFS
SEG	....
PRD	ccchhhhhhccc
MEM	....

Prosite for DKFZphfbr2_82c20.2

PS00001	8->12	ASN_GLYCOSYLATION	PDOC00001
PS00002	47->51	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	212->216	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	316->320	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005
PS00005	245->248	PKC_PHOSPHO_SITE	PDOC00005
PS00005	443->446	PKC_PHOSPHO_SITE	PDOC00005
PS00006	241->245	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00008	21->27	MYRISTYL	PDOC00008
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	28->34	MYRISTYL	PDOC00008
PS00008	48->54	MYRISTYL	PDOC00008
PS00008	231->237	MYRISTYL	PDOC00008
PS00009	2->6	AMIDATION	PDOC00009
PS00009	134->138	AMIDATION	PDOC00009
PS00029	168->190	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_82c20.2)

DKFZphfbr2_82e17
----

group: transmembrane protein

DKFZphfbr2_82e17 encodes a novel 311 amino acid protein with very weak similarity to C. elegans cosmid R01B10.

The novel protein contains 6 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans "R01B10.5" ;
membrane regions: 6
Summary DKFZphfbr2_82e17 encodes a novel 311 amino acid protein with similarity to a hypothetical C.elegans protein.

similarity to C.elegans "R01B10.5"

complete cDNA, EST HS763158 extendes the sequence, complete cds, EST hits
six potential transmembrane domains

Sequenced by DKFZ

Locus: /map="779_C_?; 818_A_1; 877_C_1; 734_C_12; 760_E_11; 171.7 cR from top of Chr14 linkage group"

Insert length: 1618 bp
Poly A stretch at pos. 1608, polyadenylation signal at pos. 1588

```
   1 CTGATCTAGT GCTTCTCGAA AAAAACCTTC AGGCGGCCCA TGGCTGTCGA
  51 TATTCAACCA GCATGCCTTG GACTTTATTG TGGGAAGACC CTATTATTTA
 101 AAAATGGCTC AACTGAAATA TATGGAGAAT GTGGGGTATG CCCAAGAGGA
 151 CAGAGAACGA ATGCACAGAA ATATTGTCAG CCTTGCACAG AATCTCCTGA
 201 ACTTTATGAT TGGCTCTATC TTGGATTTAT GGCAATGCTT CCTCTGGTTT
 251 TACATTGGTT CTTCATTGAA TGGTACTCGG GGAAAAAGAG TTCCAGCGCA
 301 CTTTTCCAAC ACATCACTGC ATTATTTGAA TGCAGCATGG CAGCTATTAT
 351 CACCTTACTT GTGAGTGATC CAGTTGGTGT TCTTTATATT CGTTCATGTC
 401 GAGTATTGAT GCTTTCTGAC TGGTACACGA TGCTTTACAA CCCAAGTCCA
 451 GATTACGTTA CCACAGTACA CTGTACTCAT GAAGCCGTCT ACCCACTATA
 501 TACCATTGTA TTTATCTATT ACGCATTCTG CTTGGTATTA ATGATGCTGC
 551 TCCGACCTCT TCTGGTGAAG AAGATTGCAT GTGGGTTAGG GAAATCTGAT
 601 CGATTTAAAA GTATTTATGC TGCACTTTAC TTCTTCCCAA TTTTAACCGT
 651 GCTTCAGGCA GTTGGTGGAG GCCTTTTATA TTACGCCTTC CCATACATTA
 701 TATTAGTGTT ATCTTTGGTT ACTCTGGCTG TGTACATGTC TGCTTCTGAA
 751 ATAGAGAACT GCTATGATCT TCTGGTCAGA AAGAAAAGAC TTATTGTTCT
 801 CTTCAGCCAC TGGTTACTTC ATGCCTATGG AATAATCTCC ATTTCCAGAG
 851 TGGATAAACT TGAGCAAGAT TTGCCCCTTT TGGCTTTGGT ACCTACACCA
 901 GCCCTTTTTT ACTTGTTCAC TGCAAAATTT ACCGAACCTT CAAGGATACT
 951 CTCAGAAGGA GCCAATGGAC ACTGAGTGTA GACATGTGAA ATGCCAAAAA
1001 CCTGAGAAGT GCTCCTAATA AAAAAGTAAA TCAATCTTAA CAGTGTATGA
1051 GAACTATTCT ATCATATATG GGAACAAGAT TGTCAGTATA TCTTAATGTT
1101 TGGGTTTGTC TTTGTTTTGT TTATGGTTAG ACTTACAGAC TTGGAAAATG
1151 CAAAACTCTG TAATACTCTG TTACACAGGG TAATATTATC TGCTACACTG
1201 GAAGGCCGCT AGGAAGCCCT TGCTTCTCTC AACAGTTCAG CTGTTCTTTA
1251 GGGCAAAATC ATGTTTCTGT GTACCTAGCA ATGTGTTCCC ATTTTATTAA
1301 GAAAAGCTTT AACACGTGTA ATCTGCAGTC CTTAACAGTG GCGTAATTGT
1351 ACGTACCTGT TGTGTTTCAG TTTGTTTTTC ACCTATAATG AATTGTAAAA
1401 ACAAACATAC TTGTGGGGTC TGATAGCAAA CATAGAAATG ATGTATATTG
1451 TTTTTTGTTA TCTATTTATT TTCATCAATA CAGTATTTTG ATGTATTGCA
1501 AAAATAGATA ATAATTTATA TAACAGGTTT TCTGTTTATA GATTGGTTCA
1551 AGATTTGTTT GGATTATTGT TCCTGTAAAG AAAACAATAA TAAAAAGCTT
1601 ACCTACATAA AAAAAAAA
```

BLAST Results

Entry HS981146 from database EMBL:
human STS WI-6253.
Length = 208
Minus Strand HSPs:
Score = 1040 (156.0 bits), Expect = 1.9e-40, P = 1.9e-40

Identities = 208/208 (100%), Positives = 208/208 (100%), Strand = Minus / Plus

Entry HSG20716 from database EMBL:
human STS A006D06.
Length = 195
Minus Strand HSPs:
Score = 975 (146.3 bits), Expect = 1.8e-37, P = 1.8e-37
Identities = 195/195 (100%), Positives = 195/195 (100%), Strand = Minus / Plus

Medline entries

No Medline entry

Peptide information for frame 1

```
  1 MAVDIQPACL GLYCGKTLLF KNGSTEIYGE CGVCPRGQRT NAQKYCQPCT
 51 ESPELYDWLY LGFMAMLPLV LHWFFIEWYS GKKSSSALFQ HITALFECSM
101 AAIITLLVSD PVGVLYIRSC RVLMLSDWYT MLYNPSPDYV TTVHCTHEAV
151 YPLYTIVFIY YAFCLVLMML LRPLLVKKIA CGLGKSDRFK SIYAALYFFP
201 ILTVLQAVGG GLLYYAFPYI ILVLSLVTLA VYMSASEIEN CYDLLVRKKR
251 LIVLFSHWLL HAYGIISISR VDKLEQDLPL LALVPTPALF YLFTAKFTEP
301 SRILSEGANG H
```

ORF from 40 bp to 972 bp; peptide length: 311
Category: similarity to unknown protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82e17, frame 1

TREMBL:AF068718_5 gene: "R01B10.5"; Caenorhabditis elegans cosmid R01B10., N = 1, Score = 399, P = 1.4e-36

>TREMBL:AF068718_5 gene: "R01B10.5"; Caenorhabditis elegans cosmid R01B10.
Length = 670

HSPs:

Score = 399 (59.9 bits), Expect = 1.4e-36, P = 1.4e-36
Identities = 95/280 (33%), Positives = 152/280 (54%)

```
Query:    2 AVDIQPACLGLYCGKTLLFKN------------GSTEIYGECGVCPRGQRTNAQKYCQPC 49
            A  IQP+CLG +CG+T+L  N             GST +   CG C  G R NA   C+ C
Sbjct:  292 ASTIQPSCLG-FCGRTVLVGNYSEDVEATTTAAGSTSL-SRCGPCSFGYRNNAMSICESC 349

Query:   50 TESPELYDWLYLGFMAMLPLVLHWFFIEWYSGKKSSSALFQ---HITALFECSMAAIITL 106
                + YDW+YL F+A+LPL+LH  FI   + K   +  ++   ++  + E  +A +I +
Sbjct:  350 DTPLQPYDWMYLLFIALLPLLLHMQFIR-IARKYCRTRYYEVSEYLCVILENVIACVIAV 408

Query:  107 LVSDPVGVLYIRSCRVLMLSDWYTMLYNPSPDYVTTVHCTHEAVYPLYTIVFIYYAFCLV 166
            L+  P     ++  C     + +WY  YNP    Y  T+ CT+E V+PLY+I FI++    +
Sbjct:  409 LIYPPRFTFFLNGCSKTDIKEWYPACYNPRIGYTKTMRCTYEVVFPLYSITFIHHLILIG 468

Query:  167 LMMLLRPLLVKKIACGLGKSDRFKSIYAALYFFPILTVLQAVGGGLLYYAFPYIILVLSL 226
              +++LR  L   +   L K+   K YAA+   PIL V+ AV  G+++Y FPYI+L+ SL
Sbjct:  469 SILVLRSTLYCVL---LYKTYNGKPFYAAIVSVPILAVIHAVLSGVVFYTFPYILLIGSL 525

Query:  227 VTLAVYMSASEIENCYDLLVR----KKRLIVLFSHWLLHAYGIISI 268
              +  +++          +++VR        LI L    L+ ++G+I+I
Sbjct:  526 WAMCFHLALEGKRPLKEMIVRIATSPTHLIFLSITMLMLSFGVIAI 571
```

Pedant information for DKFZphfbr2_82e17, frame 1

Report for DKFZphfbr2_82e17.1

[LENGTH]	311
[MW]	35239.14
[pI]	7.91
[HOMOL]	TREMBL:AF068718_5 gene: "R01B10.5"; Caenorhabditis elegans cosmid R01B10. 9e-36
[PROSITE]	AMIDATION 1
[PROSITE]	MYRISTYL 3
[PROSITE]	CAMP_PHOSPHO_SITE 1
[PROSITE]	CK2_PHOSPHO_SITE 3
[PROSITE]	PKC_PHOSPHO_SITE 4
[PROSITE]	ASN_GLYCOSYLATION 1
[KW]	TRANSMEMBRANE 6
[KW]	LOW_COMPLEXITY 7.72 %

```
SEQ   MAVDIQPACLGLYCGKTLLFKNGSTEIYGECGVCPRGQRTNAQKYCQPCTESPELYDWLY
SEG   ............................................................
PRD   cccccccccccccccceeeecccccceeeccccccccccccccceeeccccccccccchhhh
MEM   ......................................................MMMMMM

SEQ   LGFMAMLPLVLHWFFIEWYSGKKSSSALFQHITALFECSMAAIITLLVSDPVGVLYIRSC
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhccccceeeeeece
MEM   MMMMMMMMMMMMMMMM............MMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ   RVLMLSDWYTMLYNPSPDYVTTVHCTHEAVYPLYTIVFIYYAFCLVLMMLLRPLLVKKIA
SEG   ............................................xxxxxxxxxxxx....
PRD   eeeeecceeeeecccccceeeeeeeceeeeeeeeceeeeehhhhhhhhhhhhhhhhhee
MEM   ...........................MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ   CGLGKSDRFKSIYAALYFFPILTVLQAVGGGLLYYAFPYIILVLSLVTLAVYMSASEIEN
SEG   ............................................................
PRD   eecccccchhhhhhhhhhcccccccccccccceeeecceeeeehhhhhhhhhhhhhhhhhh
MEM   .........MMMMMMMMMMMMMMMMMMMMMMMMMMMMM......................

SEQ   CYDLLVRKKRLIVLFSHWLLHAYGIISISRVDKLEQDLPLLALVPTPALFYLFTAKFTEP
SEG   .....................................xxxxxxxxxxxx...........
PRD   hhhhhhhhhhhhhhhhhhhhhhhccccceeeechhhhhhceeeeeeccccceeeeeeeeccccc
MEM   .........MMMMMMMMMMMMMMMMMMMMMM.........MMMMMMMMMMMMMMMMMMMM

SEQ   SRILSEGANGH
SEG   ...........
PRD   ceeeeeccccc
MEM   MM.........
```

Prosite for DKFZphfbr2_82e17.1

PS00001	22->26	ASN_GLYCOSYLATION	PDOC00001
PS00004	82->86	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	80->83	PKC_PHOSPHO_SITE	PDOC00005
PS00005	119->122	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	294->297	PKC_PHOSPHO_SITE	PDOC00005
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00006	269->273	CK2_PHOSPHO_SITE	PDOC00006
PS00008	11->17	MYRISTYL	PDOC00008
PS00008	37->43	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00009	80->84	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_82e17.1)

DKFZphfbr2_82e4
----

group: signal transduction

DKFZphfbr2_82e4 encodes a novel 473 amino acid protein with strong similarity to the calmodulin-binding proteins.

The novel protein is similar to human and rat Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.123), rat calmodulin-binding protein, calmodulin binding protein kinase of Fugu rupies and Rattus norvegicus calcium/calmodulin-dependent protein kinase I. Calmodulin is the archetype of the family of calcium-modulated proteins of which nearly 20 members have been found. Calmodulin is involved in regulation of growth and cell cycle as well as in signal transduction and the synthesis and release of neurotransmitters. The novel protein seems to be involved in calmodulin-mediated pathways in human neuronal cells.

The new protein can find clinical application in modulating/blocking calmodulin-mediated pathways in human neuronal cells.

strong similarity to calmodulin-binding proteins

complete cDNA, complete cds, EST hits
splice variant in comparison to rat I56542
ESTs HSZZ54543/HS1141907 define splice variant
see also DKFZphfbr2_82g20 unspliced form

Sequenced by DKFZ

Locus: /map="200.5 cR from top of Chr3 linkage group"

Insert length: 2923 bp
Poly A stretch at pos. 2913, polyadenylation signal at pos. 2890

```
   1 ATGCTGGAGG TTCGCTAGCC GAAGCGGCTG CATCTGGCGC CGCGTCTGCC
  51 CCGCGTGCTC GGAGCGGATT CTGCCCGCCG TCCCCGGAGC CCTCGGCGCC
 101 CCGCTGAGCC CGCGATCACT TCCTCCCTGT GACCAACCGG CGCTGCAGGT
 151 TAGAGCCTGG CAATGCCGTT TGGGTGTGTG ACTCTGGGTG ACAAGAAGAA
 201 CTATAACCAG CCATCGGAGG TGACTGACAG ATATGATTTG GGACAGGTCA
 251 TCAAGACTGA GGAGTTTTGT GAAATCTTCC GGGCCAAGGA CAAGACGACA
 301 GGCAAGCTGC ACACCTGCAA GAAGTTCCAG AAGCGGGACG GCCGCAAGGT
 351 GCGGAAAGCT GCCAAGAACG AGATAGGCAT CCTCAAGATG GTGAAGCATC
 401 CCAACATCCT ACAGCTGGTG GATGTGTTTG TGACCCGCAA GGAGTACTTT
 451 ATCTTCCTGG AGCTGGCCAC GGGGAGGGAG GTGTTTGACT GGATCCTGGA
 501 CCAGGGCTAC TACTCGGAGC GAGACACAAG CAACGTGGTA CGGCAAGTCC
 551 TGGAGGCCGT GGCCTATTTG CACTCACTCA AGATCGTGCA CAGGAATCTC
 601 AAGCTGGAGA ACCTGGTTTA CTACAACCGG CTGAAGAACT CGAAGATTGT
 651 CATCAGTGAC TTCCATCTGG CTAAGCTAGA AAATGGCCTC ATCAAGGAGC
 701 CCTGTGGGAC CCCCGAGTAT CTGGGCAACC CACCTTTCTA TGAGGAGGTG
 751 GAAGAAGATG ATTATGAGAA CCATGATAAG AATCTCTTCC GCAAGATCCT
 801 GGCTGGTGAC TATGAGTTTG ACTCTCCATA TTGGGATGAT ATTTCGCAGG
 851 CAGCCAAAGA CCTGGTCACA AGGCTGATGG AGGTGGAGCA AGACCAGCGG
 901 ATCACTGCAG AAGAGGCCAT CTCCCATGAG TGGATTTCTG GCAATGCTGC
 951 TTCTGATAAG AACATCAAGG ATGGTGTCTG TGCCCAGATT GAAAAGAACT
1001 TTGCCAGGGC CAAGTGGAAG AAGGCTGTCC GAGTGACCAC CCTCATGAAA
1051 CGGCTCCGGG CACCAGAGCA GTCCAGCACG GCTGCAGCCC AGTCGGCCTC
1101 AGCCACAGAC ACTGCCACCC CCGGGGCTGC AGGTGGGGCC ACAGCTGCAG
1151 CTGCGAGTGG AGCTACCTCA GCCCCTGAGG GTGATGCTGC TCGTGCTGCA
1201 AAGAGTGATA ATGTGGCCCC CGCAGACCGT AGTGCCACCC CAGCCACAGA
1251 TGGAAGTGCC ACCCCAGCCA CTGATGGCAG TGTCACCCCA GCCACCGATG
1301 GAAGCATCAC TCCAGCCACT GATGGGAGTG TCACCCCAGC CACTGACAGG
1351 AGCGCTACTC CAGCCACTGA TGGGAGAGCC ACACCAGCCA CAGAAGAGAG
1401 CACTGTGCCC ACCACCCAAA GCAGTGCCAT GCTGGCCACC AAGGCAGCTG
1451 CCACCCCTGA GCCGGCTATG GCCCAGCCGG ACAGCACAGC CCCAGAGGGC
1501 GCCACAGGCC AGGCTCCACC CTCTAGTAAA GGGGAAGAGG CTGCTGGTTA
1551 TGCCCAGGAG TCTCAAAGGG AGGAGGCCAG CTGAGTAGGC AGCCTGGTGA
1601 GGGGGGGCAG GGGATGGGCA GGAGGGTGGG AGAGTGGATG AGGGGCTTCT
1651 CACTGTACAT AGAGTCACTG GCATGATGCC CTCGCTCCCC CATGCCCCCA
1701 CATCCCAGTG GGGCATAACT AGGGGTCACG GGAGAGCAGT CTCGTCTCCT
1751 GTGTGTATGT GTGTGAGTGG TGGGCAGGCC AGTGGCAGGG CCGGCCCCAG
1801 CCCCTGCATG GATTCCTTGT GGCTTTTCTG TCTTTTGCTA GCTTCACCAG
1851 TTTCTGTTCC TTGTGGGATG CTGCTCTAGG GATACTCAGG GGGCTCCTGC
1901 TCTCCTTCCC CTTCCCTTCT TGCCTCACCA TTCCCCTAGG CAGGCCCTGC
1951 AGGTCCCACA CTCTCCCAGG CCCTAAACTT GGGCGGCCTT GCCCTGAGAG
2001 CTGGTCCTCC AGCGAGGCCC TGTCAGCGGT CTTAGGCTCC TGCACATGAA
2051 GGTGTGTGCC TGTGGTGTGT GGGCTGCTCT AGGAGCAGAT ACAGGCTGGT
2101 ATAGAGGATG CAGAAAGGTA GGGCAGTATG TTTAAGTCCA GACTTGGCAC
2151 ATGGCTAGGG ATACTGCTCA CTAGCTGTGG AGGTCCTCAG GAGTGGAGAG
2201 AATGAGTAGG AGGGCAGAAG CTTCCATTTT TGTCCTTCCT AAGACCCTGT
```

```
2251 TATTTGTGTT ATTTCCTGCC TTTCCGAGTC CTGCAGTGGG CTGCCCTGTA
2301 CCCTGAACCT CATGAGCCTC TAAGGGAAAG GAGGAACAAT TAGGACGTGG
2351 CAATGAGACC TGGCAGGGCA GAGTACAAGC CCAGCACCCA GTGTCCCAGC
2401 CTTACTGGGT CCTTACCCTG GGCCAAACAG GGAGGGCTGA TACCTCCTTG
2451 CTCTTCCTAG ATGCCCACCT CCTACAATCT CAGCCCACAA GTCCTCTCCA
2501 CCCTAGGGGG CTTGCTGCAT GGCAATAACT CATAATCTGA TTTGGAGGTT
2551 TGCCCTTTAC AGGGGCAGAT TTTCTGCTCA GTTCAACAAT GAAATGAAGA
2601 GGAACTCCCT CTTTCTACAG CTCACTTCTA TCAGAGGCCC AGGTGCCTCA
2651 GAGCCACATT GAGTTGCTTT TTCTGGGATG AGGAAGTAGG GTTAAACTCC
2701 CCAGTTTCCT GAGGGAGGCT CCTGACAGGT GCCCTTTGTC AGACCCTACC
2751 ACAGCCTGGA TAGGCAGCCA CATTGGTCCT CGCCCTTGCT CGGCACTCCG
2801 TGGTGGTCCT GCCCTTCTCC CTGCATGCCT GTGGGTCTGC TCTGGTGTGT
2851 GAAGGTCGGT GGGTTAACTG TGTGCCTACT GAACCTGGCA AATAAACATC
2901 ACCCTGCAAA GCCAAAAAAA AAA
```

BLAST Results

Entry HS452352 from database EMBL:
human STS WI-15318.
Length = 350
Minus Strand HSPs:
Score = 1547 (232.1 bits), Expect = 5.2e-63, P = 5.2e-63
Identities = 331/348 (95%), Positives = 331/348 (95%), Strand = Minus / Pl

Medline entries

94110847:
J Neurosci 1994 Jan;14(1):1-13
1G5: a calmodulin-binding, vesicle-associated, protein kinase-like protein enriched in forebrain neurites.
Godbout M, Erlander MG, Hasel KW, Danielson PE, Wong KK, Battenberg EL, Foye PE,
Bloom FE, Sutcliffe JG

Peptide information for frame 1

```
  1 MPFGCVTLGD KKNYNQPSEV TDRYDLGQVI KTEEFCEIFR AKDKTTGKLH
 51 TCKKFQKRDG RKVRKAAKNE IGILKMVKHP NILQLVDVFV TRKEYFIFLE
101 LATGREVFDW ILDQGYYSER DTSNVVRQVL EAVAYLHSLK IVHRNLKLEN
151 LVYYNRLKNS KIVISDFHLA KLENGLIKEP CGTPEYLGNP PFYEEVEEDD
201 YENHDKNLFR KILAGDYEFD SPYWDDISQA AKDLVTRLME VEQDQRITAE
251 EAISHEWISG NAASDKNIKD GVCAQIEKNF ARAKWKKAVR VTTLMKRLRA
301 PEQSSTAAAQ SASATDTATP GAAGGATAAA ASGATSAPEG DAARAAKSDN
351 VAPADRSATP ATDGSATPAT DGSVTPATDG SITPATDGSV TPATDRSATP
401 ATDGRATPAT EESTVPTTQS SAMLATKAAA TPEPAMAQPD STAPEGATGQ
451 APPSSKGEEA AGYAQESQRE EAS
```

ORF from 163 bp to 1581 bp; peptide length: 473
Category: strong similarity to known protein

BLASTP hits

Entry S50193 from database PIR:
Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.123) I - rat
Length = 374
Score = 371 (130.6 bits), Expect = 2.2e-66, Sum P(2) = 2.2e-66
Identities = 74/176 (42%), Positives = 115/176 (65%)

Entry S57347 from database PIR:
Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.123) I - human
Length = 370
Score = 369 (129.9 bits), Expect = 4.6e-66, Sum P(2) = 4.6e-66
Identities = 74/176 (42%), Positives = 114/176 (64%)

Alert BLASTP hits for DKFZphfbr2_82e4, frame 1

PIR:I56542 calmodulin-binding protein - rat, N = 2, Score = 1246, P = 4e-228

TREMBLNEW:FRU010348_3 product: "calmodulin binding protein kinase"; Fugu rubripes UBE1-like gene, PRGFR2 gene and gene encoding calmodulin binding protein kinase, clone 168J21, N = 2, Score = 846, P = 2.6e-139

TREMBL:RNPRKI_1 product: "protein kinase I"; Rattus norvegicus calcium/calmodulin-dependent protein kinase I mRNA, complete cds., N = 2, Score = 364, P = 5.1e-63

```
>PIR:I56542 calmodulin-binding protein - rat
           Length = 504

  HSPs:

 Score = 1246 (186.9 bits), Expect = 4.0e-228, Sum P(2) = 4.0e-228
 Identities = 255/289 (88%), Positives = 259/289 (89%)

Query:   188 GNPPFYEEVEEDDYENHDKNLFRKILAGDYEFDSPYWDDISQAAKDLVTRLMEVEQDQRI 247
             GNPPFYEEVEEDDYENHDKNLFRKILAGDYEFDSPYWDDISQAAKDLVTRLMEVEQDQRI
Sbjct:   216 GNPPFYEEVEEDDYENHDKNLFRKILAGDYEFDSPYWDDISQAAKDLVTRLMEVEQDQRI 275

Query:   248 TAEEAISHEWISGNAASDKNIKDGVCAQIEKNFARAKWKKAVRVTTLMKRLRAPEQSSTA 307
             TAEEAISHEWISGNAASDKNIKDGVCAQIEKNFARAKWKKAVRVTTLMKRLRAPEQS TA
Sbjct:   276 TAEEAISHEWISGNAASDKNIKDGVCAQIEKNFARAKWKKAVRVTTLMKRLRAPEQSGTA 335

Query:   308 AAQSASATDTATPGAAGGATAAAASGATSAPE--------GDAARAAKSDNVAPADRSAT 359
             A      +D ATPGAAGGA AAAA GA  A          GDA  AAKSD++A ADRSAT
Sbjct:   336 AT-----SDAATPGAAGGAVAAAAGGAAPASGASATVGTGGDAGCAAKSDDMASADRSAT 390

Query:   360 PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPTTQ 419
             PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVP  Q
Sbjct:   391 PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPAAQ 450

Query:   420 SSAMLATKAAATPEPAMAQPDSTAPEGATGQAPPSSKGEEAAGYAQESQREEAS 473
             SSA  A KAAATPEPA+AQPDSTA EGATGQAPPSSKGEEA G AQESQR E S
Sbjct:   451 SSAAPAAKAAATPEPAVAQPDSTALEGATGQAPPSSKGEEATGCAQESQRVETS 504

 Score = 978 (146.7 bits), Expect = 4.0e-228, Sum P(2) = 4.0e-228
 Identities = 186/187 (99%), Positives = 187/187 (100%)

Query:     1 MPFGCVTLGDKKNYNQPSEVTDRYDLGQVIKTEEFCEIFRAKDKTTGKLHTCKKFQKRDG 60
             MPFGCVTLGDKKNYNQPSEVTDRYDLGQV+KTEEFCEIFRAKDKTTGKLHTCKKFQKRDG
Sbjct:     1 MPFGCVTLGDKKNYNQPSEVTDRYDLGQVVKTEEFCEIFRAKDKTTGKLHTCKKFQKRDG 60

Query:    61 RKVRKAAKNEIGILKMVKHPNILQLVDVFVTRKEYFIFLELATGREVFDWILDQGYYSER 120
             RKVRKAAKNEIGILKMVKHPNILQLVDVFVTRKEYFIFLELATGREVFDWILDQGYYSER
Sbjct:    61 RKVRKAAKNEIGILKMVKHPNILQLVDVFVTRKEYFIFLELATGREVFDWILDQGYYSER 120

Query:   121 DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP 180
             DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP
Sbjct:   121 DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP 180

Query:   181 CGTPEYL 187
             CGTPEYL
Sbjct:   181 CGTPEYL 187
```

Pedant information for DKFZphfbr2_82e4, frame 1

Report for DKFZphfbr2_82e4.1

[LENGTH]	473
[MW]	51208.89
[pI]	5.30
[HOMOL]	PIR:I56542 calmodulin-binding protein - rat 0.0
[FUNCAT]	30.03 organization of cytoplasm [S. cerevisiae, YFR014c] 4e-30
[FUNCAT]	10.99 other signal-transduction activities [S. cerevisiae, YFR014c] 4e-30
[FUNCAT]	03.01 cell growth [S. cerevisiae, YFR014c] 4e-30
[FUNCAT]	30.10 nuclear organization [S. cerevisiae, YKL101w] 2e-26
[FUNCAT]	03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 2e-26
[FUNCAT]	11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YDL101c] 8e-26
[FUNCAT]	98 classification not yet clear-cut [S. cerevisiae, YCL024w] 5e-24
[FUNCAT]	03.25 cytokinesis [S. cerevisiae, YDR507c] 7e-23
[FUNCAT]	03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c] 7e-23
[FUNCAT]	03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 1e-21
[FUNCAT]	03.19 recombination and dna repair [S. cerevisiae, YPL153c] 1e-21

[FUNCAT] 11.01 stress response [S. cerevisiae, YDR477w] 3e-19
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w] 3e-19
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 1e-16
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 3e-16
[FUNCAT] 03.13 meiosis [S. cerevisiae, YOR351c] 1e-15
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 3e-14
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YCR073c] 6e-11
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNR031c] 8e-11
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YJL095w] 2e-09
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 06.07 protein modification (glycolsylation, acylation, myristylation, palmitylation, farnesylation and processing) [S. cerevisiae, YFL033c] 1e-07
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YFL033c] 1e-07
[FUNCAT] 10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 5e-07
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 8e-07
[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YHR079c] 5e-06
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c] 5e-06
[FUNCAT] 30.01 organization of cell wall [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 30.90 extracellular/secretion proteins [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w] 1e-05
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae, YNL183c] 8e-05
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c] 8e-05
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDR523c] 2e-04
[FUNCAT] c energy conversion [M. genitalium, MG109] 3e-04
[BLOCKS] BL00107A Protein kinases ATP-binding region proteins
[BLOCKS] BL00939F
[SCOP] d1gol__ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus 3e-62
[SCOP] d1wfc__ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens) 5e-59
[SCOP] d1koa_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditi 1e-75
[SCOP] d1koba_ 5.1.1.1.6 Twitchin, kinase domain [california sea har 1e-72
[SCOP] d1phk__ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinas 4e-65
[SCOP] d1irk__ 5.1.1.2.4 insulin receptor [Human (Homo sapiens) 2e-56
[SCOP] d1apme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu 4e-71
[SCOP] d1fgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Hom 1e-50
[SCOP] d1ydre_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo 3e-70
[SCOP] d1fmk_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Hom 5e-49
[SCOP] d1cdkb_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su 2e-72
[SCOP] d2hcka3 5.1.1.2.1 (167-437) Haemopoetic cell kinase Hck [huma 5e-46
[SCOP] d1csn__ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe 9e-42
[SCOP] d1jsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens) 1e-56
[SCOP] d1ckia_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus) 9e-52
[EC] 2.7.1.38 Phosphorylase kinase 3e-29
[EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 8e-66
[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 2e-17
[EC] 2.7.1.117 Myosin-light-chain kinase 2e-38
[EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 2e-17
[EC] 2.7.1.37 Protein kinase 6e-28
[PIRKW] phosphotransferase 8e-66
[PIRKW] nucleus 2e-24
[PIRKW] transferase 8e-30
[PIRKW] calcium 2e-27
[PIRKW] duplication 4e-19
[PIRKW] tandem repeat 2e-31
[PIRKW] phorbol ester binding 1e-16
[PIRKW] zinc 1e-16
[PIRKW] cell cycle control 2e-20
[PIRKW] serine/threonine-specific protein kinase 8e-66
[PIRKW] phospholipid binding 1e-16
[PIRKW] autophosphorylation 8e-66
[PIRKW] brain 1e-14
[PIRKW] heterotetramer 2e-16
[PIRKW] polymer 3e-29
[PIRKW] mitosis 2e-20
[PIRKW] magnesium 7e-22
[PIRKW] ATP 8e-66
[PIRKW] alternative initiators 1e-29

```
[PIRKW]         phosphoprotein 8e-66
[PIRKW]         apoptosis 2e-31
[PIRKW]         glycoprotein 4e-19
[PIRKW]         skeletal muscle 3e-28
[PIRKW]         protein kinase 2e-28
[PIRKW]         testis 3e-28
[PIRKW]         signal transduction 1e-21
[PIRKW]         cAMP binding 1e-16
[PIRKW]         purine nucleotide binding 5e-25
[PIRKW]         structural protein 4e-19
[PIRKW]         calcium binding 3e-45
[PIRKW]         alternative splicing 3e-45
[PIRKW]         P-loop 5e-25
[PIRKW]         lipoprotein 2e-16
[PIRKW]         cardiac muscle 4e-19
[PIRKW]         muscle 3e-28
[PIRKW]         myristylation 2e-16
[PIRKW]         EF hand 5e-29
[PIRKW]         cell division 2e-38
[PIRKW]         calmodulin binding 8e-66
[PIRKW]         smooth muscle 7e-31
[SUPFAM]        fibronectin type III repeat homology 7e-31
[SUPFAM]        immunoglobulin homology 7e-31
[SUPFAM]        ribosomal protein S6 kinase II 3e-26
[SUPFAM]        calcium-dependent protein kinase 5e-29
[SUPFAM]        AMP-activated protein kinase 7e-22
[SUPFAM]        protein kinase akt 1e-14
[SUPFAM]        protein kinase SPK1 3e-20
[SUPFAM]        unassigned Ser/Thr or Tyr-specific protein kinases 2e-36
[SUPFAM]        Ca2+/calmodulin-dependent protein kinase 3e-45
[SUPFAM]        calmodulin repeat homology 5e-29
[SUPFAM]        protein kinase DUN1 2e-24
[SUPFAM]        Dictyostelium cAMP-dependent protein kinase catalytic chain 1e-14
[SUPFAM]        death-associated protein kinase 2e-31
[SUPFAM]        myosin-light-chain kinase, nonmuscle 1e-29
[SUPFAM]        pleckstrin repeat homology 1e-14
[SUPFAM]        ankyrin repeat homology 2e-31
[SUPFAM]        protein kinase homology 8e-66
[SUPFAM]        Ca2+/calmodulin-dependent protein kinase II 8e-36
[SUPFAM]        twitchin 1e-18
[SUPFAM]        protein kinase C zinc-binding repeat homology 1e-16
[SUPFAM]        titin 4e-19
[SUPFAM]        protein kinase cdr1 2e-20
[SUPFAM]        kinase-related transforming protein 2e-38
[SUPFAM]        Ca2+/calmodulin-dependent protein kinase I 8e-66
[SUPFAM]        kinase interaction domain homology 2e-24
[SUPFAM]        protein kinase C mu 1e-16
[PROSITE]       AMIDATION       1
[PROSITE]       MYRISTYL        3
[PROSITE]       CK2_PHOSPHO_SITE        10
[PROSITE]       TYR_PHOSPHO_SITE        2
[PROSITE]       PKC_PHOSPHO_SITE        11
[PFAM]          Eukaryotic protein kinase domain
[KW]            All_Alpha
[KW]            3D
[KW]            LOW_COMPLEXITY      7.40 %

SEQ     MPFGCVTLGDKKNYNQPSEVTDRYDLGQVIKTEEFCEIFRAKDKTTGKLHTCKKFQKRDG
SEG     ............................................................
1a06-   .............CEETTTGGGCEEEEEEECBCGGGGGEEEEEETTTTCEEEEEEEEEC---

SEQ     RKVRKAAKNEIGILKMVKHPNILQLVDVFVTRKEYFIFLELATGREVFDWILDQGYYSER
SEG     ............................................................
1a06-   -------HHHHHHHHHCCTTTBCCEEEEEEEETTEEEEEEECCCCCEEHHHHHHHTTTTBHH

SEQ     DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP
SEG     ............................................................
1a06-   HHHHHHHHHHHHHHHHHHHCCCTTTTTTTTEEECCCTTTTCEEECCCTTTTCHHHHHCCC

SEQ     CGTPEYLGNPPFYEEVEEDDYENHDKNLFRKILAGDYEFDSPYWDDISQAAKDLVTRLME
SEG     ............................................................
1a06-   HHHHHHHCCTTTTTT--------THHHHHHHHHCCCCCCTTTTTTTTCHHHHHHHHHHCT

SEQ     VEQDQRITAEEAISHEWISGNAASDKNIKDGVCAQIEKNFARAKWKKAVRVTTLMKRLRA
SEG     ............................................................
1a06-   TTGGGCCCHHHHHHTTTTTTCCCCCCBHHHHHHHHHHHHHHCCTTTTTTBTHHHHHHHC..

SEQ     PEQSSTAAAQSASATDTATPGAAGGATAAAASGATSAPEGDAARAAKSDNVAPADRSATP
SEG     ..xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx......................
1a06-   ............................................................
```

```
SEQ     ATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPTTQS
SEG     ............................................................
1a06-   ............................................................

SEQ     SAMLATKAAATPEPAMAQPDSTAPEGATGQAPPSSKGEEAAGYAQESQREEAS
SEG     .....................................................
1a06-   .....................................................
```

Prosite for DKFZphfbr2_82e4.1

PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	51->54	PKC_PHOSPHO_SITE	PDOC00005
PS00005	91->94	PKC_PHOSPHO_SITE	PDOC00005
PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	118->121	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	264->267	PKC_PHOSPHO_SITE	PDOC00005
PS00005	394->397	PKC_PHOSPHO_SITE	PDOC00005
PS00005	454->457	PKC_PHOSPHO_SITE	PDOC00005
PS00005	467->470	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	91->95	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	118->122	CK2_PHOSPHO_SITE	PDOC00006
PS00006	248->252	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	336->340	CK2_PHOSPHO_SITE	PDOC00006
PS00006	442->446	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	467->471	CK2_PHOSPHO_SITE	PDOC00006
PS00007	456->464	TYR_PHOSPHO_SITE	PDOC00007
PS00007	127->136	TYR_PHOSPHO_SITE	PDOC00007
PS00008	260->266	MYRISTYL	PDOC00008
PS00008	321->327	MYRISTYL	PDOC00008
PS00008	324->330	MYRISTYL	PDOC00008
PS00009	59->63	AMIDATION	PDOC00009

Pfam for DKFZphfbr2_82e4.1

```
HMM_NAME      Eukaryotic protein kinase domain

HMM                *YeigRiIGeGsFGtVYkCiWr.TGeIVAIKIIkkrsms.....FlREIq
                    Y +G++I     F ++++++++ TG++    K++ KR+         + +EI
Query          24   YDLGQVIKTEEFCEIFRAKDKTTGKLHTCKKFQKRDGRKVRKAAKNEIG     72

HMM                IMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEwe
                   I+++++HPNI+++ D+F  + +++ + +E++ G + FD+I ++G++SE++
Query          73  ILKMVKHPNILQLVDVFV-TRKEYFIFLELATGREVFDWILDQGYYSERD   121

HMM                IrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeN...gqIKIcDFGLAR
                     ++++Q+L++++YLHS +I+HR LK EN+ + ++      I I+DF LA+
Query         122  TSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAK   171

HMM                qMnnYerMttfCGTPWY*
                   + N  ++ +  CGTP+Y
Query         172  LEN--GLIKEPCGTPEY    186

HMM                *GepPFyd...........dnMemImrIiqrfrrpfWpnCSeElyDFMr
                    G PPFY+            + +++I++++++F +P+W+ +S  ++D+++
Query         188   GNPPFYEEVEEDDYENHDKNLFRKILAGDYEFDSPYWDDISQAAKDLVT   236

HMM                wCWnyDPekRPTFrQILnHPWF*
                   +++++  ++R+T+++++ H W+
Query         237  RLMEVEQDQRITAEEAISHEWI    258
```

DKFZphfbr2_82g14
----

group: transmembrane protein

DKFZphfbr2_82g14 encodes a novel 208 amino acid proline-rich protein without similarity to known proteins.

The protein contains one transmembrane domain.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

unknown prolin rich protein
membrane regions: 1
Summary DKFZphfbr2_82g14 encodes a novel 208 amino acid protein.

unknown prolin rich protein

complete cDNA, complete cds, EST hits
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="26.2 cR from top of Chr16 linkage group"

Insert length: 2059 bp
Poly A stretch at pos. 2049, polyadenylation signal at pos. 2024

```
   1 AGAAGTGCGA CTGCCAGCTG CCGAGGCGTT CGGTCCTGCT GTTGCGGCCG
  51 CTGCCCCAGG GCTGCGGGGA CGCTCCCGGA GCCCTGCCTG TCCCCTGTCC
 101 ATCCAGGCCA GCAGCTGAAG GAGCCTCACC TGCCTCCCTT CTCTGAGTAG
 151 CACGGATTTG AGGAGAAGCA GCGAAGATGT CCAGCGAGCC TCCCCCTCCT
 201 TATCCTGGGG GCCCCACAGC CCCACTTCTG GAAGAGAAAA GTGGAGCCCC
 251 GCCCACCCCA GGCCGTTCCT CCCCAGCTGT GATGCAGCCC CCTCCAGGCA
 301 TGCCACTGCC CCCTGCGGAC ATTGGCCCCC CACCCTATGA GCCGCCGGGT
 351 CACCCAATGC CCCAGCCTGG CTTCATCCCA CCACACATGA GTGCAGATGG
 401 CACCTACATG CCTCCGGGTT TCTACCCTCC TCCAGGCCCC CACCCACCCA
 451 TGGGCTACTA CCCCCCAGGG CCCTACACGC CAGGGCCCTA CCCTGGCCCT
 501 GGGGGCCACA CAGCCACAGT CCTGGTCCCT TCAGGAGCTG CCACCACGGT
 551 GACAGTGCTG CAGGGAGAGA TCTTTGAGGG AGCGCCTGTG CAGACGGTGT
 601 GTCCCCACTG CCAGCAGGCC ATCGCCACCA AGATCTCCTA CGAGATTGGC
 651 TTGATGAATT TCGTGCTGGG TTTCTTCTGT TGCTTCATGG GATGTGATCT
 701 GGGCTGCTGC CTGATCCCCT GCCTCATCAA TGACTTCAAG GATGTGACGC
 751 ACACATGCCC CAGCTGCAAA GCCTACATCT ACACGTACAA GCGCCTGTGC
 801 TAACGGAGCT GGGACTCGGG ACTCCCCCGC CTGTCAGTCT GGCCCCCTGT
 851 GCTTTGCTCC CTGCGCTCAG TGGTCACTTT CCCGCTCCCA CTTGGGGCTG
 901 GGAGCCGTGC CACCATCCCC TAGAAGTCCT GTCCTCTTCA CCCTGCCCTA
 951 CCTGAGCCGC TGACTCTTCT GGCAAAAATT CTGTTGGGAT TTAAGGCCAA
1001 GGGTCAGTGG GTGGCAGGGG GCTGGCAATG AGCTTGTGTG TTGTTGGTCT
1051 GCTTGGTGTG TGTGATCGGG AAGATAAGCT GGGAGGGGTC TCCTGCTGGG
1101 GTCCTGATGC CTCTGTTTCC AAACAAGGTA CAGGTTCAGT CCAGACTCTT
1151 TCCCCCTGGG ACCAACAGCA GCCAGAGCAG TTAGCCAGTT AGTCCCCAGG
1201 CCTGTGGCCA CAGGCGTTTC TGACCTGCTG GGCCGAGAAT GGGTAAGTTG
1251 TCTGGAGTCA GGTGGGCCCA CGTAGGACAG GGTCACAAAG CCTGGGTTTG
1301 TTTCTGGGTA CTTTGCGCCT CTGGGGTGCT AGAGGTGGGG CATGGTGGCT
1351 GGAAGTAAAA CTGCCAACTC TGGCCCTCAG AACTCTCAGG TATAGAAGCC
1401 CAGGATGTCT AATACCCTGT CCCAGTGCCC GAGAGCTGCC TGGTGTCAGG
1451 TAGAGAGGAC ACTGTACCTG GGTGAATGAT CAGACCCTGG TAGCTAAGAA
1501 GGAACTTGTC CCTTTGAGTC AGTGTGCAGA CCCCCTTTCA GGCCATGCCT
1551 CTGTGAACCC TGTATTGCTG GGGCCGGAAG GAGCCCCTGA GCCTAGCCCC
1601 TTCCCGTCTG CCCTGTGTCC TCACTGCGTG TGGGTATGAC CTCTGCCTGG
1651 TGGCTGGTGT ATCCCAACTG GGCAAGAGAT GGCAGAGGGT CCCCCTTGTG
1701 GGTGCGCTTG GATGTGCAGA GCCTTCTCCA TGGATTTTCT TCCCTGTAAG
1751 TGCCGGGCCC CCCACCCCAG CTGACAGGCT GTTGCTGTGC CTGCTCACAC
1801 CTGCTCCTGC AGGCACACTG GGCTAGGGAC GAGGAAGGAG CAGCCACAAG
1851 TGGTAGAACT GCCTTGGTGG ACACCAGCCT CGCCCTGTCT TTATTTCCTG
1901 AATGGTTTGT GAACTTGCTC ACCTGGACCA CTGTATCCTG CCACTGTCCT
1951 TCCTGGTCTC GCACTGCCAC TGCATGGCCT CCTGTCACTG TGAATCGTGG
2001 CCCAGTCTCA GTTTGTAGTT TCTCATTAAA TTGGCCCTTT CACTCCCCCA
2051 AAAAAAAAA
```

BLAST Results

Entry HS727347 from database EMBL:
human STS WI-16589.
Length = 275
Plus Strand HSPs:
Score = 1365 (204.8 bits), Expect = 3.0e-55, P = 3.0e-55
Identities = 275/276 (99%), Positives = 275/276 (99%), Strand = Plus /
Pl

Medline entries

No Medline entry

Peptide information for frame 3

```
  1 MSSEPPPPYP GGPTAPLLEE KSGAPPTPGR SSPAVMQPPP GMPLPPADIG
 51 PPPYEPPGHP MPQPGFIPPH MSADGTYMPP GFYPPPGPHP PMGYYPPGPY
101 TPGPYPGPGG HTATVLVPSG AATTVTVLQG EIFEGAPVQT VCPHCQQAIA
151 TKISYEIGLM NFVLGFFCCF MGCDLGCCLI PCLINDFKDV THTCPSCKAY
201 IYTYKRLC
```

ORF from 177 bp to 800 bp; peptide length: 208
Category: similarity to known protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82g14, frame 3

PIR:S57447 HPBRII-7 protein - human, N = 1, Score = 206, P = 8.4e-16

PIR:A47655 spliceosome-associated protein SAP 62 - human, N = 1, Score = 198, P = 4.3e-15

>PIR:S57447 HPBRII-7 protein - human
Length = 551

HSPs:

Score = 206 (30.9 bits), Expect = 8.4e-16, P = 8.4e-16
Identities = 57/115 (49%), Positives = 62/115 (53%)

```
Query:    5 PPPPYPGGPTAPLLEEKSGAPPTPGRSSPAVMQPPPGMPLPPADIGPP-----PYEP--- 56
            PPPP+P G T P       G P PG   P    PPPG LPP    GPP     P  P
Sbjct:  226 PPPPFPAGQTPP--RPPLGPPGPPGPPGP----PPPGQVLPPPLAGPPNRGDRPPPPVLF 279

Query:   57 PGHPMPQP--GFIPPHMSADGTYMP-PGFYPPPGPHPPM-GYYPP-GPYTPGPYPGPGGH 111
            PG P  QP  G +PP      G    P PG+ PPPGP PP  G  PP GP+ P P PGP G
Sbjct:  280 PGQPFGQPPLGPLPP-----GPPPPVPGYGPPPGPPPPQQGPPPPPGPFPPRP-PGPLGP 333

Query:  112 TATVLVP 118
              T+  P
Sbjct:  334 PLTLAPP 340
```

Score = 177 (26.6 bits), Expect = 1.1e-12, P = 1.1e-12
Identities = 55/120 (45%), Positives = 61/120 (50%)

```
Query:    5 PPPPYPGGPTAP--LLEEKSGAPPTPG-RSSPAVM---QP---PPGMPLPPADIGPPPYE 55
            P PP P GP  P  +L       PP  G R  P V+   QP   PP  PLPP    GPPP
Sbjct:  244 PGPPGPPGPPPPGQVLPPPLAGPPNRGDRPPPPVLFPGQPFGQPPLGPLPP---GPPP-P 299

Query:   56 PPGHPMPQPGFIPPHMSADGTYMPPGFYPP--PGP-HPPMGYYPPGPYTPGPYPG---PG 109
            PG+  P PG  PP      G    PPG +PP  PGP  PP+    PP P+ PGP PG   P
Sbjct:  300 VPGYG-PPPGPPPPQQ---GPPPPPGPFPPRPPGPLGPPLTLAPP-PHLPGPPPGAPPPA 354

Query:  110 GHTATVLVP 118
             H     P
Sbjct:  355 PHVNPAFFP 363
```

Score = 168 (25.2 bits), Expect = 1.1e-11, P = 1.1e-11
Identities = 47/118 (39%), Positives = 51/118 (43%)

```
Query:    5 PPPPYPG-GPTAPLLEEKSGAPPTPGRSSPAVMQP--PPGMPLPPADI-GPPPYEPPGHP 60
```

```
                PPPP PG GP       + G PP PG    P    P  PP     PP  + GPPP   PP  P
Sbjct:   296 PPPPVPGYGPPPGPPPPQQGPPPPPGPFPPRPPGPLGPPLTLAPPPHLPGPPPGAPPPAP 355

Query:    61 MPQPGFIPPHMSADGTYMPPGFYPPPGPHPPMGYYPPGPYTPGPYPGPGGHTATVLVPSG 120
                P F PP  ++      MP      P P  P G  PP PY  G Y  PG     T   P
Sbjct:   356 HVNPAFFPPPTNSG---MPTSDSRGPPPTDPYGR-PP-PYDRGDYGPPGREMDTARTPLS 410

Query:   121 AA 122
              A
Sbjct:   411 EA 412

 Score = 156 (23.4 bits), Expect = 2.1e-10, P = 2.1e-10
 Identities = 44/103 (42%), Positives = 50/103 (48%)

Query:     6 PPPYPGGPTAPLLEEKSGAPPT-PGRSSPAVMQPPPGMPLPPADIGPPPYEPPGHPMPQP 64
             P   PGG    P         G PP  P    +P   +PP G P PP    GPPP    PG  +P P
Sbjct:   208 PGAVPGGDRFPGPAGPGGPPPPFPAGQTPP--RPPLGPPGPPGPPGPPP---PGQVLPPP 262

Query:    65 GFIPPHMSADGTYMPPGFYP-PPGPHPPMGYYPPGPYTP----GPYPGP 108
                 PP+    D     PP  +P  P    PP+G  PPGP  P      GP PGP
Sbjct:   263 LAGPPNRG-DRP-PPPVLFPGQPFGQPPLGPLPPGPPPPVPGYGPPPGP 309

 Score = 121 (18.2 bits), Expect = 5.2e-05, P = 5.2e-05
 Identities = 40/90 (44%), Positives = 45/90 (50%)

Query:    23 GAPPTPGRSSPAVMQPP-PGMPLPPAD-IGPP-PYEPPGHPMPQPG-FIPPHMSADGTYM 78
             G     PG + P       PP P     PP     +GPP P  PPG P P PG  +PP ++
Sbjct:   213 GGDRFPGPAGPGGPPPPFPAGQTPPRPPLGPPGPPGPPG-P-PPPGQVLPPPLAG----- 265

Query:    79 PP--GFYPPPG---PHPPMGYYPPGPYTPGPYPG-PG 109
             PP   G   PPP    P  P G  P GP   PGP P  PG
Sbjct:   266 PPNRGDRPPPPVLFPGQPFGQPPLGPLPPGPPPPVPG 302
```

Pedant information for DKFZphfbr2_82g14, frame 3

Report for DKFZphfbr2_82g14.3

```
[LENGTH]    208
[MW]        21862.47
[pI]        5.55
[PROSITE]   MYRISTYL        3
[PROSITE]   PKC_PHOSPHO_SITE    2
[KW]        TRANSMEMBRANE  1
[KW]        LOW_COMPLEXITY    39.90 %
```

```
SEQ   MSSEPPPPYPGGPTAPLLEEKSGAPPTPGRSSPAVMQPPPGMPLPPADIGPPPYEPPGHP
SEG   ....xxxxxxxxx...................xxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD   cccccccccccccchhhhhhcccccccccccccccccccccccccccccccccccccccc
MEM   ............................................................

SEQ   MPQPGFIPPHMSADGTYMPPGFYPPPGPHPPMGYYPPGPYTPGPYPGPGGHTATVLVPSG
SEG   xxxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.........
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccceeeeeecccc
MEM   ............................................................

SEQ   AATTVTVLQGEIFEGAPVQTVCPHCQQAIATKISYEIGLMNFVLGFFCCFMGCDLGCCLI
SEG   ............................................................
PRD   cceeeeeeeeeeeccccceeeecccchhhhhhhhhhhhhhceeeeeeeeeeeccccceeec
MEM   ...............................................MMMMMMMMMMMMM

SEQ   PCLINDFKDVTHTCPSCKAYIYTYKRLC
SEG   ............................
PRD   eeeecccccccccccccceeeeeeeccc
MEM   MMMM........................
```

Prosite for DKFZphfbr2_82g14.3

PS00005	196->199	PKC_PHOSPHO_SITE	PDOC00005
PS00005	203->206	PKC_PHOSPHO_SITE	PDOC00005
PS00008	109->115	MYRISTYL	PDOC00008
PS00008	120->126	MYRISTYL	PDOC00008
PS00008	172->178	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_82g14.3)

DKFZphfbr2_82i17
----

group: signal transduction

DKFZphtes2_82i17 encodes a novel 334 amino acid protein with similarity to the plasma membrane substrate for the cAMP-dependent protein kinase.

The novel protein is a transmembrane protein with strong similarity to the phospholemman protein, a membrane substrate for the cAMP-dependent protein kinase. It seems to serve as a chloride channel or as a chloride-channel regulator.

The new protein can find application in modulating/blocking cAMP-dependent protein kinase-dependent pathways.

similarity to plasma membrane substrate for cAMP-dependent protein kinase

complete cDNA, complete cds, EST hits
potential start at Bp 31 matches Kozak consensus PyNNatgG
might be a SODIUM/POTASSIUM-TRANSPORTING ATPASE
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="11; 920_E_12; 786_(A,H)_11; (797,802)_(E,H)_7"

Insert length: 1647 bp
Poly A stretch at pos. 1637, polyadenylation signal at pos. 1615

```
   1 AGTCTCGGAG GGGACCGGCT GTGCAGACGC CATGGAGTTG GTGCTGGTCT
  51 TCCTCTGCAG CCTGCTGGCC CCCATGGTCC TGGCCAGTGC AGCTGAAAAG
 101 GAGAAGGAAA TGGACCCTTT TCATTATGAT TACCAGACCC TGAGGATTGG
 151 GGGACTGGTG TTCGCTGTGG TTCTCTTCTC GGTTGGGATC CTCCTTATCC
 201 TAAGTCGCAG GTGCAAGTGC AGTTTCAATC AGAAGCCCCG GGCCCCAGGA
 251 GATGAGGAAG CCCAGGTGGA GAACCTCATC ACCGCCAATG CAACAGAGCC
 301 CCAGAAAGCA GAGAACTGAA GTGCAGCCAT CAGGTGGAAG CCTCTGGAAC
 351 CTGAGGCGGC TGCTTGAACC TTTGGATGCA AATGTCGATG CTTAAGAAAA
 401 CCGGCCACTT CAGCAACAGC CCTTTCCCCA GGAGAAGCCA AGAACTTGTG
 451 TGTCCCCCAC CCTATCCCCT CTAACACCAT TCCTCCACCT GATGATGCAA
 501 CTAACACTTG CCTCCCCGCT GCAGCCTGTG GTCCTGCCCA CCTCCCGTGA
 551 TGTGTGTGTG TGTGTGTGTG TGTGTGACTG TGTGTGTTTG CTAACTGTGG
 601 TCTTTGTGGC TACTTGTTTG TGGATGGTAT TGTGTTTGTT AGTGAACTGT
 651 GGACTCGCTT TCCCAGGCAG GGGCTGAGCC ACACGGCCAT CTGCTCCTCC
 701 CTGCCCCCGT GGCCCTCCAT CACCTTCTGC TCCTAGGAGG CTGCTTGTTG
 751 CCCGAGACCA GCCCCCTCCC CTGATTTAGG GATGCGTAGG GTAAGAGCAC
 801 GGGCAGTGGT CTTCAGTCGT CTTGGGACCT GGGAAGGTTT GCAGCACTTT
 851 GTCATCATTC TTCATGGACT CCTTTCACTC CTTTAACAAA AACCTTGCTT
 901 CCTTATCCCA CCTGATCCCA GTCTGAAGGT CTCTTAGCAA CTGGAGATAC
 951 AAAGCAAGGA GCTGGTGAGC CCAGCGTTGA CGTCAGGCAG GCTATGCCCT
1001 TCCGTGGTTA ATTTCTTCCC AGGGGCTTCC ACGAGGAGTC CCCATCTGCC
1051 CCGCCCCTTC ACAGAGCGCC CGGGGATTCC AGGCCCAGGG CTTCTACTCT
1101 GCCCCTGGGG AATGTGTCCC CTGCATATCT TCTCAGCAAT AACTCCATGG
1151 GCTCTGGGAC CCTACCCCTT CCAACCTTCC CTGCTTCTGA GACTTCAATC
1201 TACAGCCCAG CTCATCCAGA TGCAGACTAC AGTCCCTGCA ATTGGGTCTC
1251 TGGCAGGCAA TAGTTGAAGG ACTTCCTGTT CCGTTGGGGC CAGCACACCG
1301 GGATGGATGG AGGGAGAGCA GAGGCCTTTG CTTCTCTGCC TACGTCCCCT
1351 TAGATGGGCA GCAGAGGCAA CTCCCGCATC CTTTGCTCTG CCTGTCAGTG
1401 GTCAGAGCGG TGAGCGAGGT GGGTTGGAGA CTCAGCAGGC TCCGTGCAGC
1451 CCTTGGGAAC AGTGAGAGGT TGAAGGTCAT AACGAGAGTG GGAACTCAAC
1501 CCAGATCCCG CCCCTCCTGT CCTCTGTGTT CCCGCGGAAA CCAACCAAAC
1551 CGTGCGCTGT GACCCATTGC TGTTCTCTGT ATCGTGACCT ATCCTCAACA
1601 ACAACAGAAA AAAGGAATAA AATATCCTTT GTTTCCTAAA AAAAAAA
```

BLAST Results

Entry HS31455 from database EMBL:
human STS WI-2739.
Length = 103
Minus Strand HSPs:
Score = 487 (73.1 bits), Expect = 4.4e-14, P = 4.4e-14
Identities = 101/104 (97%), Positives = 101/104 (97%), Strand = Minus / Plus
frame shift in primer binding site

Medline entries

91250422:
Purification and complete sequence determination of the major plasma membrane substrate
for cAMP-dependent protein kinase and protein kinase C in myocardium.

95091702:
Protein kinase C and cyclic AMP-dependent protein kinase phosphorylate phospholemman,
an insulin and adrenaline-regulated membrane phosphoprotein, at specific sites in the
carboxy terminal domain.

95138184:
Mat-8, a novel phospholemman-like protein expressed in human breast tumors, induces a
chloride conductance in Xenopus oocytes.

Peptide information for frame 2

```
 1 MELVLVFLCS LLAPMVLASA AEKEKEMDPF HYDYQTLRIG GLVFAVVLFS
51 VGILLILSRR CKCSFNQKPR APGDEEAQVE NLITANATEP QKAEN
```

ORF from 32 bp to 316 bp; peptide length: 95
Category: strong similarity to known protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82i17, frame 2

SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR., N = 1, Score = 196, P = 1.2e-15

TREMBL:AF091390_1 product: "phospholemman precursor"; Mus musculus phospholemman precursor, gene, complete cds., N = 1, Score = 187, P = 1.1e-14

PIR:A40533 cAMP-dependent protein kinase major membrane substrate precursor - dog, N = 1, Score = 189, P = 6.5e-15

SWISSPROT:PLM_RAT PHOSPHOLEMMAN PRECURSOR., N = 1, Score = 185, P = 1.7e-14

```
>SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR.
            Length = 92

  HSPs:

 Score = 196 (29.4 bits), Expect = 1.2e-15, P = 1.2e-15
 Identities = 43/85 (50%), Positives = 56/85 (65%)

Query:     4 VLVFLCSLLAPMVLASAAEKEKEMDPFHYDYQTLRIGGLVFAVVLFSVGILLILSRRCKC 63
             +LVF   LL      + AE  KE DPF YDYQ+L+IGGLV  A +LF +GIL++LSRRC+C
Sbjct:     7 ILVFCVGLLT----MAKAESPKEHDPFTYDYQSLQIGGLVIAGILFILGILIVLSRRCRC 62

Query:    64 SFNQKPRA--PGDEEAQVENLITANAT 88
              FNQ+ R   P +EE    + I    +T
Sbjct:    63 KFNQQQRTGEPDEEEGTFRSSIRRLST 89
```

Pedant information for DKFZphfbr2_82i17, frame 2

Report for DKFZphfbr2_82i17.2

[LENGTH]	95
[MW]	10542.37
[pI]	5.05
[HOMOL]	SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR. 3e-15
[BLOCKS]	BL01310

[EC]	3.6.1.37 Na+/K+-exchanging ATPase 6e-08
[PIRKW]	transmembrane protein 1e-09
[PIRKW]	hydrolase 6e-08
[PROSITE]	ATP1G1_PLM_MAT8 1
[PROSITE]	MYRISTYL 1
[PROSITE]	CK2_PHOSPHO_SITE 1
[PROSITE]	TYR_PHOSPHO_SITE 1
[PROSITE]	PKC_PHOSPHO_SITE 2
[PROSITE]	ASN_GLYCOSYLATION 1
[KW]	Alpha_Beta
[KW]	SIGNAL_PEPTIDE 19

SEQ MELVLVFLCSLLAPMVLASAAEKEKEMDPFHYDYQTLRIGGLVFAVVLFSVGILLILSRR
PRD ccchhhhhhhhhhccccccccccccccccccceeeeecccceeeehhhhhhheeeeehhh

SEQ CKCSFNQKPRAPGDEEAQVENLITANATEPQKAEN
PRD hhhccccccccccchhhhhhhhhhhccccccccc

Prosite for DKFZphfbr2_82i17.2

PS00001	86->90	ASN_GLYCOSYLATION	PDOC00001
PS00005	36->39	PKC_PHOSPHO_SITE	PDOC00005
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS00007	25->33	TYR_PHOSPHO_SITE	PDOC00007
PS00008	41->47	MYRISTYL	PDOC00008
PS01310	28->42	ATP1G1_PLM_MAT8	PDOC01014

(No Pfam data available for DKFZphfbr2_82i17.2)

DKFZphfbr2_82i24
----

group: nucleic acid management

DKFZphfbr2_82i24 encodes a novel 547 amino acid protein with similarity to DEAD-box superfamily ATP-dependent helicases.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis.
The novel protein contains a DEAD-box an ATP/GTP-binding site motif A (P-loop, interacting with one of the phophate groups of the nucleotide) and a leucine zipper. Mutations in the closely related Drosophila Hlc gene result in lethality in homozygotes. Therefore the new protein seems to be critical involved in RNA processing in eukariontic c ells.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to DEAD-box subfamily ATP-dependent helicase

complete cDNA, complete cds, EST hits
potential Start at Bp 9 matches Kozak consensus PyNNatgG,
[PFAM] Helicases conserved C-terminal domain
[PFAM] DEAD and DEAH box helicases

Sequenced by DKFZ

Locus: /map="720_A_3; 758_H_4; 772_E_3; 804_A_5; 175.5 cR from topFT of Chr7 linkage group"

Insert length: 1860 bp
Poly A stretch at pos. 1850, polyadenylation signal at pos. 1829

```
   1 AGCAGCGCCA TGGAGGACTC TGAAGCACTG GGCTTCGAAC ACATGGGCCT
  51 CGATCCCCGG CTCCTTCAGG CTGTCACCGA TCTGGGCTGG TCGCGACCTA
 101 CGCTGATCCA GGAGAAGGCC ATCCCACTGG CCCTAGAAGG GAAGGACCTC
 151 CTGGCTCGGG CCCGCACGGG CTCCGGGAAG ACGGCCGCTT ATGCTATTCC
 201 GATGCTGCAG CTGTTGCTCC ATAGGAAGGC GACAGGTCCG GTGGTAGAAC
 251 AGGCAGTGAG AGGCCTTGTT CTTGTTCCTA CCAAGGAGCT GGCACGGCAA
 301 GCACAGTCCA TGATTCAGCA GCTGGCTACC TACTGTGCTC GGGATGTCCG
 351 AGTGGCCAAT GTCTCAGCTG CTGAAGACTC AGTCTCTCAG AGAGCTGTGC
 401 TGATGGAGAA GCCAGATGTG GTAGTAGGGA CCCCATCTCG CATATTAAGC
 451 CACTTGCAGC AAGACAGCCT GAAACTTCGT GACTCCCTGG AGCTTTTGGT
 501 GGTGGACGAA GCTGACCTTC TTTTTTCCTT TGGCTTTGAA GAAGAGCTCA
 551 AGAGTCTCCT CTGTCACTTG CCCCGGATTT ACCAGGCTTT TCTCATGTCA
 601 GCTACTTTTA ACGAGGACGT ACAAGCACTC AAGGAGCTGA TATTACATAA
 651 CCCGGTTACC CTTAAGTTAC AGGAGTCCCA GCTGCCTGGG CCAGACCAGT
 701 TACAGCAGTT TCAGGTGGTC TGTGAGACTG AGGAAGACAA ATTCCTCCTG
 751 CTGTATGCCC TGCTCAAGCT GTCATTGATT CGGGGCAAGT CTCTGCTCTT
 801 TGTCAACACT CTAGAACGGA GTTACCGGCT ACGCCTGTTC TTGGAACAGT
 851 TCAGCATCCC CACCTGTGTG CTCAATGGAG AGCTTCCACT GCGCTCCAGG
 901 TGCCACATCA TCTCACAGTT CAACCAAGGC TTCTACGACT GTGTCATAGC
 951 AACTGATGCT GAAGTCCTGG GGGCCCCAGT CAAGGGCAAG CGTCGGGGCC
1001 GAGGGCCCAA AGGGGACAAG GCCTCTGATC CGGAAGCAGG TGTGGCCCGG
1051 GGCATAGACT TCCACCATGT GTCTGCTGTG CTCAACTTTG ATCTTCCCCC
1101 AACCCCTGAG GCCTACATCC ATCGAGCTGG CAGGACAGCA CGCGCTAACA
1151 ACCCAGGCAT AGTCTTAACC TTTGTGCTTC CCACGGAGCA GTTCCACTTA
1201 GGCAAGATTG AGGAGCTTCT CAGTGGAGAG AACAGGGGCC CCATTCTGCT
1251 CCCCTACCAG TTCCGGATGG AGGAGATCGA GGGCTTCCGC TATCGCTGCA
1301 GGGATGCCAT GCGCTCAGTG ACTAAGCAGG CCATTCGGGA GGCAAGATTG
1351 AAGGAGATCA AGGAAGAGCT TCTGCATTCT GAGAAGCTTA AGACATACTT
1401 TGAAGACAAC CCTAGGGACC TCCAGCTGCT GCGGCATGAC CTACCTTTGC
1451 ACCCCGCAGT GGTGAAGCCC CACCTGGGCC ATGTTCCTGA CTACCTGGTT
1501 CCTCCTGCTC TCCGTGGCCT GGTACGCCCT CACAAGAAGC GGAAGAAGCT
1551 GTCTTCCTCT TGTAGGAAGG CCAAGAGAGC AAAGTCCCAG AACCCACTGC
1601 GCAGCTTCAA GCACAAAGGA AAGAAATTCA GACCCACAGC CAAGCCCTCC
1651 TGAGGTTGTT GGGCCTCTCT GGAGCTGAGC ACATTGTGGA GCACAGGCTT
1701 ACACCCTTCG TGGACAGGCG AGGCTCTGGT GCTTACTGCA CAGCCTGAAC
1751 AGACAGTTCT GGGGCCGGCA GTGCTGGGCC CTTTAGCTCC TTGGCACTTC
1801 CAAGCTGGCA TCTTGCCCCT TGACAACAGA ATAAAAATTT TAGCTGCCCC
1851 AAAAAAAAAA
```

BLAST Results

Entry HSG05793 from database EMBL:
human STS WI-6581.
Length = 206
Minus Strand HSPs:
Score = 992 (148.8 bits), Expect = 6.0e-38, P = 6.0e-38
Identities = 204/208 (98%), Positives = 204/208 (98%), Strand = Minus / P1

Entry AC004938 from database EMBL:
Homo sapiens clone DJ0971C03; HTGS phase 1, 18 unordered pieces.
Score = 1269, P = 6.5e-202, identities = 269/282
12 exons. Bp ~87920-93706 (matching 1-1497)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 10 bp to 1650 bp; peptide length: 547
Category: strong similarity to known protein
Classification: Nucleic acid management
Prosite motifs: ATP_GTP_A (51-59)
LEUCINE_ZIPPER (149-171)

```
  1 MEDSEALGFE HMGLDPRLLQ AVTDLGWSRP TLIQEKAIPL ALEGKDLLAR
 51 ARTGSGKTAA YAIPMLQLLL HRKATGPVVE QAVRGLVLVP TKELARQAQS
101 MIQQLATYCA RDVRVANVSA AEDSVSQRAV LMEKPDVVVG TPSRILSHLQ
151 QDSLKLRDSL ELLVVDEADL LFSFGFEEEL KSLLCHLPRI YQAFLMSATF
201 NEDVQALKEL ILHNPVTLKL QESQLPGPDQ LQQFQVVCET EEDKFLLLYA
251 LLKLSLIRGK SLLFVNTLER SYRLRLFLEQ FSIPTCVLNG ELPLRSRCHI
301 ISQFNQGFYD CVIATDAEVL GAPVKGKRRG RGPKGDKASD PEAGVARGID
351 FHHVSAVLNF DLPPTPEAYI HRAGRTARAN NPGIVLTFVL PTEQFHLGKI
401 EELLSGENRG PILLPYQFRM EEIEGFRYRC RDAMRSVTKQ AIREARLKEI
451 KEELLHSEKL KTYFEDNPRD LQLLRHDLPL HPAVVKPHLG HVPDYLVPPA
501 LRGLVRPHKK RKKLSSSCRK AKRAKSQNPL RSFKHKGKKF RPTAKPS
```

BLASTP hits

No BLASTP hits available
Alert BLASTP hits for DKFZphfbr2_82i24, frame 1

TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (iby), waclaw (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst), and la costa (lcs) genes, complete cds., N = 1, Score = 1230, P = 3.2e-125

TREMBL:SPCC1494_6 gene: "SPCC1494.06c"; product: "atp dependent helicase"; S.pombe chromosome II cosmid c1494., N = 2, Score = 753, P = 2.5e-113

PIR:S51412 hypothetical protein YLR276c - yeast (Saccharomyces cerevisiae), N = 2, Score = 711, P = 8.2e-117

TREMBL:AF025451_2 gene: "C24H12.4"; Caenorhabditis elegans cosmid C24H12., N = 2, Score = 564, P = 2.7e-99

>TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (iby), waclaw (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst), and la costa (lcs) genes, complete cds.
Length = 560

HSPs:

Score = 1230 (184.5 bits), Expect = 3.2e-125, P = 3.2e-125
Identities = 251/497 (50%), Positives = 344/497 (69%)

```
Query:    9 FEHMGLDPRLLQAVTDLGWSRPTLIQEKAIPLALEGKDLLARARTGSGKTAAYAIPMLQL 68
            F  + LD R+L+AV  LGW +PTLIQ  AIPL LEGKD++ RARTGSGKTA YA+P++Q
Sbjct:   11 FHELELDQRILKAVAQLGWQQPTLIQSTAIPLLLEGKDVVVRARTGSGKTATYALPLIQK 70

Query:   69 LLHRKATGPVVEQAVRGLVLVPTKELARQAQSMIQQLATYCARDVRVANVS-AAEDSVSQ 127
            +L+ K        EQ V  +VL PTKEL RQ++ +I+QL   C + VRVA+++ ++ D+V+Q
Sbjct:   71 ILNSKLNAS--EQYVSAVVLAPTKELCRQSRKVIEQLVESCGKVVRVADIADSSNDTVTQ 128

Query:  128 RAVLMEKPDVVVGTPSRILSHLQQDSLKLRDSLELLVVDEADLLFSFGFEEELKSLLCHL 187
            R  L E PD+VV TP+ +L++ +  S+      +E LVVDEADL+F++G+E++ K L+ HL
Sbjct:  129 RHALSESPDIVVATPANLLAYAEAGSVVDLKHVETLVVDEADLVFAYGYEKDFKRLIKHL 188

Query:  188 PRIYQAFLMSATFNEDVQALKELILHNPVTLKLQESQLPGPDQLQQFQVVCETEEDKFLL 247
            P IYQA L+SAT  +DV  +K L L+NPVTLKL+E +L   DQL   +++ E E DK  +
Sbjct:  189 PPIYQAVLVSATLTDDVVRMKGLCLNNPVTLKLEEPELVPQDQLSHQRILAE-ENDKPAI 247

Query:  248 LYALLKLSLIRGKSLLFVNTLERSYRLRLFLEQFSIPTCVLNGELPLRSRCHIISQFNQG 307
            LYALLKL LIRGKS++FVN+++R Y++RLFLEQF I  CVLN ELP   R H ISQFN+G
Sbjct:  248 LYALLKLRLIRGKSIIFVNSIDRCYKVRLFLEQFGIRACVLNSELPANIRIHTISQFNKG 307

Query:  308 FYDCVIATDAEVLGAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNFDLPPTPE 367
            YD +IA+D   +  P  G +      K  ++ D E+  +RGIDF  V+ V+NFD P
Sbjct:  308 TYDIIIASDEHHMEKP--GGKSATNRKSPRSGDMESSASRGIDFQCVNNVINFDFPRDVT 365

Query:  368 AYIHRAGRTARANNPGIVLTFVLPTEQFHLGKIEELL----SGENRGPILLPYQFRMEEI 423
            +YIHRAGRTAR NN G VL+FV   E       +E+ L    + +     I+  YQF+MEE+
Sbjct:  366 SYIHRAGRTARGNNKGSVLSFVSMKESKVNDSVEKKLCDSFAAQEGEQIIKNYQFKMEEV 425

Query:  424 EGFRYRCRDAMRSVTKQAIREARLKEIKEELLHSEKLKTYFEDNPRDLQLLRHDLPLHPA 483
            E FRYR +D  R+ T+ A+ + R++EIK E+L+ EKLK +FE+N RDLQ LRHD PL
Sbjct:  426 ESFRYRAQDCWRAATRVAVHDTRIREIKIEILNCEKLKAFFEENKRDLQALRHDKPLRAI 485

Query:  484 VVKPHLGHVPDYLVPPALRGLV 505
            V+ HL  +P+Y+VP AL+ +V
Sbjct:  486 KVQSHLSDMPEYIVPKALKRVV 507
```

Pedant information for DKFZphfbr2_82i24, frame 1

Report for DKFZphfbr2_82i24.1

[LENGTH] 547
[MW] 61589.88
[pI] 9.34
[HOMOL] TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (iby), waclaw (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst), and la costa (lcs) genes, complete cds. 1e-121
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YLR276c] 1e-109
[FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 2e-42
[FUNCAT] 04.01.04 rrna processing [S. cerevisiae, YLL008w] 8e-40
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 8e-40
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YLL008w] 8e-40
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YKR059w] 3e-39
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YKR059w] 3e-39
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YDL160c] 3e-35
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPL119c] 3e-29
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 4e-29
[FUNCAT] 1 genome replication, transcription, recombination and repair [H. influenzae, HI0892] 1e-27
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 2e-27
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 4e-21
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL064c] 1e-05
[BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW] nucleus 4e-34
[PIRKW] RNA binding 7e-41
[PIRKW] DEAD box 2e-38
[PIRKW] transmembrane protein 9e-20
[PIRKW] DNA binding 8e-23
[PIRKW] ATP 1e-107
[PIRKW] purine nucleotide binding 2e-38
[PIRKW] P-loop 1e-107
[PIRKW] hydrolase 2e-35
[PIRKW] protein biosynthesis 2e-38
[PIRKW] ATP binding 7e-43

```
[SUPFAM]        WW repeat homology 1e-26
[SUPFAM]        DEAD/H box helicase homology 1e-107
[SUPFAM]        unassigned DEAD/H box helicases 1e-107
[SUPFAM]        ATP-dependent RNA helicase DBP1 3e-31
[SUPFAM]        ATP-dependent RNA helicase DHH1 2e-35
[SUPFAM]        translation initiation factor eIF-4A 2e-38
[SUPFAM]        tobacco ATP-dependent RNA helicase DB10 1e-26
[PROSITE]       ATP_GTP_A      1
[PROSITE]       LEUCINE_ZIPPER  1
[PFAM]          Helicases conserved C-terminal domain
[PFAM]          DEAD and DEAH box helicases
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY      9.87 %

SEQ     MEDSEALGFEHMGLDPRLLQAVTDLGWSRPTLIQEKAIPLALEGKDLLARARTGSGKTAA
SEG     ............................................................
PRD     cccccccccccccchhhhhhhhhcccccccccccccccccccccceeeeecccccccccee

SEQ     YAIPMLQLLLHRKATGPVVEQAVRGLVLVPTKELARQAQSMIQQLATYCARDVRVANVSA
SEG     ............................................................
PRD     ehhhhhhhhhhcccccccccceeeeeeccchhhhhhhhhhhhhhhhhhcceeeeeecc

SEQ     AEDSVSQRAVLMEKPDVVVGTPSRILSHLQQDSLKLRDSLELLVVDEADLLFSFGFEEEL
SEG     .......................................xxxxxxxxxxxx.........
PRD     ccchhhhhhhhhccccceeeecccccchhhhhhccccchhhhhhhhhhhhhhhhcchhhh

SEQ     KSLLCHLPRIYQAFLMSATFNEDVQALKELILHNPVTLKLQESQLPGPDQLQQFQVVCET
SEG     ............................................................
PRD     hhhhhhccchhhhhhhhhcccchhhhhhhhhhhccccceeeeecccccchhhhhhhhhhh

SEQ     EEDKFLLLYALLKLSLIRGKSLLFVNTLERSYRLRLFLEQFSIPTCVLNGELPLRSRCHI
SEG     .....xxxxxxxxxxx............................................
PRD     hhhhhhhhhhhhhhhhhccceeeeeeehhhhhhhhhhhhhccccceeeccccchhhhhhhh

SEQ     ISQFNQGFYDCVIATDAEVLGAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNF
SEG     .........................xxxxxxxxxxxxx......................
PRD     hhhhhccceeeeeeccccccccccccccccccccccccccccccccccccccceeeeeec

SEQ     DLPPTPEAYIHRAGRTARANNPGIVLTFVLPTEQFHLGKIEELLSGENRGPILLPYQFRM
SEG     ............................................................
PRD     cccccccceeeecccccccccccccccceeeeeeecchhhhhhhhhhhhhhhccccccccccchh

SEQ     EEIEGFRYRCRDAMRSVTKQAIREARLKEIKEELLHSEKLKTYFEDNPRDLQLLRHDLPL
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhccc

SEQ     HPAVVKPHLGHVPDYLVPPALRGLVRPHKKRKKLSSSCRKAKRAKSQNPLRSFKHKGKKF
SEG     ............................xxxxxxxxxxxxxxxxxx..............
PRD     ccccccccccccceeeccccccccccccccccccccchhhhhhccccccccccccccccc

SEQ     RPTAKPS
SEG     .......
PRD     ccccccc
```

Prosite for DKFZphfbr2_82i24.1

```
PS00017       51->59    ATP_GTP_A           PDOC00017
PS00029      149->171   LEUCINE_ZIPPER      PDOC00029
```

Pfam for DKFZphfbr2_82i24.1

```
HMM_NAME        DEAD and DEAH box helicases

HMM                  *gLpPWILRnIyeMGFEkPTPIQQqAIPiILeGRDVMACAQTGSGKTAAF
                      GL+P +L +++++G+++PT IQ++AIP++LEG+D++A+A TGSGKTAA+
Query            13   GLDPRLLQAVTDLGWSRPTLIQEKAIPLALEGKDLLARARTGSGKTAAY    61

HMM                  lIPMLQHIDwdP...WpqpPQdPrALILAPTRELAMQIQEEcRkFgkHMn
                     +IPMLQ +++ +    +  +   +R+L+L+PT ELA+Q Q  +++++ ++
Query            62  AIPMLQLLLHRKATGPVVEQA-VRGLVLVPTKELARQAQSMIQQLATYCA   110

HMM                  g.IRImcIYGGtnMRdQMRmLeRGpPHIVIATPGRLIDHIERgtldLDr.
                       +R++ +     +   Q  +L+++P ++V++TP R++ H+++ +L+L++
Query           111  RDVRVANVSAAEDSVSQRAVLMEKP-DVVVGTPSRILSHLQQDSLKLRDS   159

HMM                  IeMLVMDEADRMLDMGFIDQIRrIMrqIPMpwNRQTMMFSATMPdeIqEL
                     +E LV DEAD +++ GF++++  ++ ++P  +  Q  + SAT+ +++Q L
```

```
Query        160 LELLVVDEADLLFSFGFEEELKSLLCHLP--RIYQAFLMSATFNEDVQAL   207

HMM              ARrFMRNPIRInIdMdElTtnEnIkQwYiyVerEMWKfdcLcrLIe*
                  + +++NP+ + + +++L  +  ++Q+ +++E E++KF +L+ L++
Query        208 KELILHNPVTLKLQESQLPGPDQLQQFQVVCETEEDKFLLLYALLK   253


HMM_NAME          Helicases conserved C-terminal domain

HMM              *EileeWLknlGIrvmYIHGdMpQeERdeIMddFNnGEynVLIcTDV...
                   +L+ +L++  I+++++ G +P + R  I+ +FN+G Y++ I+TD+
Query        272  YRLRLFLEQFSIPTCVLNGELPLRSRCHIISQFNQGFYDCVIATDAEVL   320

HMM              .......................ggRGIDIPdVNHVINYDMPWNPEqYI
                                         +RGID+  V+ V N+D+P +PE YI
Query        321 GAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNFDLPPTPEAYI   370

HMM              QRIGRTgRIG*
                 +R+GRT+R++
Query        371 HRAGRTARAN     380
```

DKFZphfbr2_82m16
----

group: brain derived

DKFZphfbr2_82m16 encodes a novel 289 amino acid protein with very weak similarity to A.thaliana F28A23.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A.thaliana F28A23.140

complete cDNA, complete cds, few EST hits
many ATGs in front of the ORF
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="4"

Insert length: 2715 bp
Poly A stretch at pos. 2705, polyadenylation signal at pos. 2687

```
   1 AGAGGAGGGG AGAGGACTGG GGAGCCGAGC CAGAGCCGGG CTGCCTGCCA
  51 CCCGGCTGCT CGTCCGCTAG CTGGGGAGGA GCGCTCCACC CGCAACTGAC
 101 AAAGGATGGG AGAATGCCCG CGCCCCGGGA TGCCGGCCGC ACGCAGCCTG
 151 GCGGCCGCCT GAGCTACTTC ACCCTCCGCC GGTAAGTGAC TGCAAACATC
 201 ATTCATTCAA TCAGCCTCAC TGGGAGCCCC TTCTCTCCGG CTGGTAGTCC
 251 TGGGCGGCTT GTCCCTGATC CCGAGCGGGG CTTGGCACAG CATCAGCCCT
 301 GGAGGGCAGG CAGCAGGTGC CTTTGCCTGG TGGGTCCACT GGGGAGCGTG
 351 GCTGGGGTTC GCGGCGGGTG CTGCCACCCA ACCTGCGGGC GGCGGGCTCG
 401 CCCAGTAGGC GCCTCTCTGG TGAGAGGAGG CGGCTCCAGC CCGCATCCTG
 451 GGGTAGTTGC TACTATTGGC CCCCAGCGCC CGCTCTGCGC GCGCGCCGTT
 501 TCTGGCGGAT CCCCAGTGCG CGGCGCGCTG TTTACACCGG CGTGGTACTA
 551 GTCACGGAGC CGCACCCCTC GGAAAGCGCG GAGTCGATGA CAGCCACTTC
 601 ACAGGCTCAC GCGCTCCTAG TGTGGGCTTG AAGGGGACGG GGACCGATTA
 651 CCAAAGGAGA GCGCTGAGTA CGGAAGACAC AGGGCAGCCT TTGTCTTGGG
 701 TTTAGCGCTG ATGCGCTCAA CCCTGAGTCG GGTTCACTGC AACTGTTGTG
 751 TCCGATTTCG GTTCCCTGCA ACCGCCCTCC TGGGCGAGAG ATGTCATTGT
 801 GTTCCTGCGG CCAGCGGGAC TGAGAGCTGG GACTTAAGAC GCCAGGAGGG
 851 TCCTGCGCTC ACGGGAAATG TACCCCAAAA GAACTCTGAG AGAATATACT
 901 CAACTGTCCT GCTGTGATTA AACAAGACTG CTGTATTTTA ATTTCAGAAA
 951 TTGAAAAGGG ATAGGAGGAA GGGGAAAATG CTGGGCTGGT GTGAAGCGAT
1001 AGCCCGTAAC CCTCACAGAA TTCCAAACAA CACGCGAACA CCCGAGATCT
1051 CAGGGGATTT GGCTGACGCC TCACAAACCT CCACATTGAA TGAAAAATCC
1101 CCAGGGCGAT CTGCAAGTCG ATCAAGTAAC ATTTCAAAAG CAAGCAGCCC
1151 AACAACAGGG ACAGCTCCCA GGAGCCAGTC AAGGTTGTCT GTCTGTCCAT
1201 CCACTCAGGA CATCTGCAGA ATCTGTCACT GCGAAGGGGA TGAAGAGAGC
1251 CCCCTCATCA CACCCTGTCG CTGCACTGGG ACACTGCGCT TTGTCCACCA
1301 GTCCTGCCTC CACCAGTGGA TAAAGAGCTC AGATACACGC TGCTGTGAGC
1351 TCTGCAAGTA TGACTTCATA ATGGAGACCA AGCTCAAACC CCTCCGGAAG
1401 TGGGAGAAAC TACAGATGAC CACAAGTGAA AGGAGGAAAA TATTCTGCTC
1451 TGTCACATTC CACGTAATCG CGATCACCTG TGTGGTTTGG TCTTTGTATG
1501 TATTGATAGA CCGGACAGCG GAGGAAATCA AGCAAGGCAA TGACAATGGT
1551 GTCCTTGAAT GGCCATTTTG GACAAAACTG GTTGTGGTAG CCATTGGCTT
1601 CACAGGAGGT CTTGTCTTCA TGTACGTACA GTGTAAAGTC TATGTTCAGT
1651 TGTGGCGCAG GCTGAAGGCC TACAACCGTG TGATCTTTGT ACAAAATTGC
1701 CCAGACACTG CCAAAAAACT GGAGAAGAAC TTCTCATGTA ATGTAAACAC
1751 AGACATCAAA GATGCTGTGG TAGTGCCTGT ACCACAAACA GGTGCAAATT
1801 CACTGCCATC TGCAGAGGGT GGCCCCCCTG AAGTTGTATC AGTCTGATGG
1851 AACCTGTTGG GAGTTTCTTC ACCGAAGAAT ATCTTTCTAG CCCTCAGCCA
1901 CTACAAATGA CAGAAGTGAC CTTGAATTAT TTACTCCCTT CAGCTCCTCC
1951 TTTCTCCTAC TGACACATTT TTCCTGACTT TGTTCAAAGA GGAAAGGAGA
2001 AAAACAAACA AACAGACCAA ATGCCCAGGA GCCCATGAAG TAATAGCGTA
2051 AAGTAAAGTA TGATATGGAA ATGTGAAGTT TGCAAGAGAA TGATTTCCAA
2101 GACAATTAAG AACTACTGGG GCAATGAATG CTTTTAGGCA GTAATCAAAG
2151 ATTAAATGGA CCCATGATAC TCTTCTTCAC AGTAACAGGG GAAAAGTTCA
2201 AGAATACAGA CTTGAATTGC GATGTGTATT ACTTCTAGGG CCTTGTAATG
2251 TTAACTGTCT CATCTGGAAA TAATAACTAA CATATTTGGT TTTAAGCCTG
2301 AAATTGTCTG CATTATCCCT AAGTCACATT GGAAGTGAAC TTGGAGGATG
2351 CATATTTTGA TATGCTTTGA CAGCTAACAG ATTTGTATGG TTTAGTGGAG
2401 TCTGGTTATT TTGACAGATG CATGTTTTTT TTAAATAGAT GCAATATACA
2451 TTTGAAGACA TTGATATTTG GAATTAATTA TGTTTGTTTA AGTCACGCAA
2501 AAGATTTTCA GAAAATGTTC GGATATAATT AGCTCTGTTA AATACCCACA
2551 GAACTGTTAT CAGGTCTTAT ATTTATTTTC ATCTGGTTCC TCTAATACAG
```

```
2601 TGCTGTCCAA TAGAAACACA ACAGCCACAA ATGCAGGCCA CAGATGCAAA
2651 TATTTAACTT CCCAGTAGCC CTATTTTAAA AAGTAAAAAT AAATGTTTGT
2701 TTGTTAAAAA AAAAA
```

BLAST Results

Entry G37457 from database EMBLNEW:
SHGC-57357 Human Homo sapiens STS genomic.
Length = 458
Plus Strand HSPs:
Score = 2116 (317.5 bits), Expect = 4.3e-91, P = 4.3e-91
Identities = 444/456 (97%)

Medline entries

No Medline entry

Peptide information for frame 3

```
  1 MLGWCEAIAR NPHRIPNNTR TPEISGDLAD ASQTSTLNEK SPGRSASRSS
 51 NISKASSPTT GTAPRSQSRL SVCPSTQDIC RICHCEGDEE SPLITPCRCT
101 GTLRFVHQSC LHQWIKSSDT RCCELCKYDF IMETKLKPLR KWEKLQMTTS
151 ERRKIFCSVT FHVIAITCVV WSLYVLIDRT AEEIKQGNDN GVLEWPFWTK
201 LVVVAIGFTG GLVFMYVQCK VYVQLWRRLK AYNRVIFVQN CPDTAKKLEK
251 NFSCNVNTDI KDAVVVPVPQ TGANSLPSAE GGPPEVVSV
```

ORF from 978 bp to 1844 bp; peptide length: 289
Category: similarity to unknown protein

BLASTP hits

Entry AB011169_1 from database TREMBL:
gene: "KIAA0597"; product: "KIAA0597 protein"; Homo sapiens mRNA for KIAA0597 protein, partial cds.
Score = 188, P = 6.0e-12, identities = 30/54, positives = 38/54

Entry SPBC14F5_7 from database TREMBL:
gene: "SPBC14F5.07"; product: "hypothetical protein"; S.pombe chromosome II cosmid c14F5.
Score = 185, P = 1.9e-11, identities = 29/53, positives = 38/53

Entry CEY57A10B_1 from database TREMBL:
gene: "Y57A10B.1"; Caenorhabditis elegans cosmid Y57A10B
Score = 171, P = 2.6e-10, identities = 40/107, positives = 58/107

Alert BLASTP hits for DKFZphfbr2_82m16, frame 3

TREMBL:ATF28A23_14 gene: "F28A23.140"; product: "putative protein"; Arabidopsis thaliana DNA chromosome 4, BAC clone F28A23 (ESSAII project), N = 1, Score = 198, P = 3.4e-13

>TREMBL:ATF28A23_14 gene: "F28A23.140"; product: "putative protein";
Arabidopsis thaliana DNA chromosome 4, BAC clone F28A23 (ESSAII project)
Length = 1,051

HSPs:

Score = 198 (29.7 bits), Expect = 3.4e-13, P = 3.4e-13
Identities = 38/103 (36%), Positives = 61/103 (59%)

```
Query:   28 LADASQTSTLNEKSPGRSASRS-SNISKASSPTTGTAPRSQSRLSVCPSTQDICRICHCE 86
            +++ S +S+ +  SP +++    SN+  A S  TG+               +D+CRIC
Sbjct:   20 VSEPSVSSSSSSSSPNQASPNPFSNMDPAVSTATGSRYVDDDE-----DEEDVCRICRNP 74

Query:   87 GDEESPLITPCRCTGTLRFVHQSCLHQWIKSSDTRCCELCKYDF 130
            GD ++PL  PC C+G+++FVHQ CL QW+  S+ R CE+CK+ F
Sbjct:   75 GDADNPLRYPCACSGSIKFVHQDCLLQWLNHSNARQCEVCKHPF 118
```

Pedant information for DKFZphfbr2_82m16, frame 3

Report for DKFZphfbr2_82m16.3

[LENGTH]	289
[MW]	32308.36
[pI]	8.76
[HOMOL]	PIR:T00268 hypothetical protein KIAA0597 - human (fragment) 9e-14
[FUNCAT]	04.99 other transcription activities [S. cerevisiae, YIL030c] 4e-09
[PIRKW]	transmembrane protein 9e-08
[PROSITE]	MYRISTYL 1
[PROSITE]	CK2_PHOSPHO_SITE 4
[PROSITE]	TYR_PHOSPHO_SITE 1
[PROSITE]	PKC_PHOSPHO_SITE 3
[PROSITE]	ASN_GLYCOSYLATION 3
[KW]	Alpha_Beta
[KW]	LOW_COMPLEXITY 6.57 %

```
SEQ   MLGWCEAIARNPHRIPNNTRTPEISGDLADASQTSTLNEKSPGRSASRSSNISKASSPTT
SEG   ..................................xxxxxxxxxxxxxxxxxxx..
PRD   ccchhhhhhccccccccccccccccchhhhhhhhhcccccccccccccccccccccccc

SEQ   GTAPRSQSRLSVCPSTQDICRICHCEGDEESPLITPCRCTGTLRFVHQSCLHQWIKSSDT
SEG   ............................................................
PRD   cccccccccccccccccceeeeeeeccccccccccccccccccceeeeehhhhhhhhcccc

SEQ   RCCELCKYDFIMETKLKPLRKWEKLQMTTSERRKIFCSVTFHVIAITCVVWSLYVLIDRT
SEG   ............................................................
PRD   ceeeeeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc

SEQ   AEEIKQGNDNGVLEWPFWTKLVVVAIGFTGGLVFMYVQCKVYVQLWRRLKAYNRVIFVQN
SEG   ............................................................
PRD   cccccccccccceeehhhhheeeeeeecccccceeeeehhhhhhhhhhhhhhheeeeeee

SEQ   CPDTAKKLEKNFSCNVNTDIKDAVVVPVPQTGANSLPSAEGGPPEVVSV
SEG   .................................................
PRD   ccchhhhhhccccccccccceeeeeeeccccccccccccccccccccc
```

Prosite for DKFZphfbr2_82m16.3

PS00001	17->21	ASN_GLYCOSYLATION	PDOC00001
PS00001	51->55	ASN_GLYCOSYLATION	PDOC00001
PS00001	251->255	ASN_GLYCOSYLATION	PDOC00001
PS00005	102->105	PKC_PHOSPHO_SITE	PDOC00005
PS00005	150->153	PKC_PHOSPHO_SITE	PDOC00005
PS00005	244->247	PKC_PHOSPHO_SITE	PDOC00005
PS00006	36->40	CK2_PHOSPHO_SITE	PDOC00006
PS00006	75->79	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00007	121->129	TYR_PHOSPHO_SITE	PDOC00007
PS00008	187->193	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_82m16.3)

DKFZphfbr2_82m6
----

group: signal transduction

DKFZphfbr2_82m6.3 encodes a novel 654 amino acid protein with similarity to murine sphingosine kinase.

Sphingosine kinase is a new type of lipid kinase, which is regulated by growth factors. The enzyme phosphorylates sphingosine, which subsequently exerts intracellular and extracellular actions. Intracellulary, sphingosine 1-phosphate (SPP) promotes proliferation and inhibits apoptosis. In yeast, survival of cells exposed to heat shock indicates is dependend on SPP. Extracellulary, SPP inhibits cell motility and influences cell morphology, effects that appear to be mediated by the G protein-coupled receptor EDG1.

The new protein can find application in modulating/blocking the shingosine kinase intracellular signal transmission pathway.

strong similarity to mouse "sphingosine kinase"

complete cDNA, complete cds, EST hits,
YLR260w/YOR171c Lcb5p/Lcb4p = long chain base kinases,
involved in biosynthesis of sphingolipids

Sequenced by DKFZ

Locus: unknown

Insert length: 2875 bp
Poly A stretch at pos. 2865, polyadenylation signal at pos. 2838

```
   1 AGTGTTGGAG GTGAGGAGGC GGGGCTGGCA GGGCTAGTCG GGGCATCTGG
  51 AAATTTCCGA CCCCACGCTT CGGGCGTTTC CTTATCAGGT TCACCGCTCC
 101 CTGATCTCGC GCTGCACTTC GTAGGCGCAG CCGCTGCTTG GGAAGTCCTA
 151 CTTAAGAGCT GAAGGTCAGG CCAGGACAGT GAGACCTGAC TCCTTGCTCC
 201 TACCAGCCTA CTATGGCTTA AGACCCAGGG CCAGGGTCCC GTTGATGTAA
 251 CAGAGCAGAG GACCAGCAGA TGAATGGACA CCTTGAAGCA GAGGAGCAGC
 301 AGGACCAGAG GCCAGACCAG GAGCTGACCG GGAGCTGGGG CCACGGGCCT
 351 AGGAGCACCC TGGTCAGGGC TAAGGCCATG GCCCCGCCCC CACCGCCACT
 401 GGCTGCCAGC ACCTCGCTCC TCCATGGCGA GTTTGGCTCC TACCCAGCCC
 451 GAGGCCCACG CTTTGCCCTC ACCCTTACAT CGCAGGCCCT GCACATACAG
 501 CGGCTGCGCC CCAAACCTGA AGCCAGGCCC CGGGGTGGCC TGGTCCCGTT
 551 GGCCGAGGTC TCAGGCTGCT GCACCCTGCG AAGCCGCAGC CCCTCAGACT
 601 CAGCGGCCTA CTTCTGCATC TACACCTACC CTCGGGGCCG GCGCGGGGCC
 651 CGGCGCAGAG CCACTCGCAC CTTCCGGGCA GATGGGGCCG CCACCTACGA
 701 AGAGAACCGT GCCGAGGCCC AGCGCTGGGC CACTGCCCTC ACCTGTCTGC
 751 TCCGAGGACT GCCACTGCCC GGGGATGGGG AGATCACCCC TGACCTGCTA
 801 CCTCGGCCGC CCCGGTTGCT TCTATTGGTC AATCCCTTTG GGGGTCGGGG
 851 CCTGGCCTGG CAGTGGTGTA AGAACCACGT GCTTCCCATG ATCTCTGAAG
 901 CTGGGCTGTC CTTCAACCTC ATCCAGACAG AACGACAGAA CCACGCCCGG
 951 GAGCTGGTCC AGGGGCTGAG CCTGAGTGAG TGGGATGGCA TCGTCACGGT
1001 CTCGGGAGAC GGGCTGCTCC ATGAGGTGCT GAACGGGCTC CTAGATCGCC
1051 CTGACTGGGA GGAAGCTGTG AAGATGCCTG TGGGCATCCT CCCCTGCGGC
1101 TCGGGCAACG CGCTGGCCGG AGCAGTGAAC CAGCACGGGG GATTTGAGCC
1151 AGCCCTGGGC CTCGACCTGT TGCTCAACTG CTCACTGTTG CTGTGCCGGG
1201 GTGGTGGCCA CCCACTGGAC CTGCTCTCCG TGACGCTGGC CTCGGGCTCC
1251 CGCTGTTTCT CCTTCCTGTC TGTGGCCTGG GGCTTCGTGT CAGATGTGGA
1301 TATCCAGAGC GAGCGCTTCA GGGCCTTGGG CAGTGCCCGC TTCACACTGG
1351 GCACGGTGCT GGGCCTCGCC ACACTGCACA CCTACCGCGG ACGCCTCTCC
1401 TACCTCCCCG CCACTGTGGA ACCTGCCTCG CCCACCCCTG CCCATAGCCT
1451 GCCTCGTGCC AAGTCGGAGC TGACCCTAAC CCCAGACCCA GCCCCGCCCA
1501 TGGCCCACTC ACCCCTGCAT CGTTCTGTGT CTGACCTGCC TCTTCCCCTG
1551 CCCCAGCCTG CCCTGGCCTC TCCTGGCTCG CCAGAACCCC TGCCCATCCT
1601 GTCCCTCAAC GGTGGGGGCC CAGAGCTGGC TGGGGACTGG GGTGGGGCTG
1651 GGGATGCTCC GCTGTCCCCG GACCCACTGC TGTCTTCACC TCCTGGCTCT
1701 CCCAAGGCAG CTCTACACTC ACCCGTCTCC GAAGGGGCCC CCGTAATTCC
1751 CCCATCCTCT GGGCTCCCAC TTCCCACCCC TGATGCCCGG GTAGGGGCCT
1801 CCACCTGCGG CCCGCCCGAC CACCTGCTGC CTCCGCTAGG CACCCCGCTG
1851 CCCCCAGACT GGGTGACGCT GGAGGGGGAC TTTGTGCTCA TGTTGGCCAT
1901 CTCGCCCAGC CACCTAGGCG CTGACCTGGT GGCAGCTCCG CATGCGCGCT
1951 TCGACGACGG CCTGGTGCAC CTGTGCTGGG TGCGTAGCGG CATCTCGCGG
2001 GCTGCGCTGC TGCGCCTTTT CTTGGCCATG GAGCGTGGTA GCCACTTCAG
2051 CCTGGGCTGT CCGCAGCTGG GCTACGCCGC GGCCCGTGCC TTCCGCCTAG
2101 AGCCGCTCAC ACCACGCGGC GTGCTCACAG TGGACGGGGA GCAGGTGGAG
2151 TATGGGCCGC TACAGGCACA GATGCACCCT GGCATCGGTA CACTGCTCAC
2201 TGGGCCTCCT GGCTGCCCGG GGCGGGAGCC CTGAAACTAA ACAAGCTTGG
2251 TACCCGCCGG GGGCGGGGCC TACATTCCAA TGGGGCGGAG CCTGAGCTAG
2301 GGGGTGTGGC CTGGCTGCTA GAGTTGTGGT GGCAGGGGCC CTGGCCCCGT
```

```
2351 CTCAGGATTG CGCTCGCTTT CATGGGACCA GACGTGATGC TGGAAGGTGG
2401 GCGTCGTCAC GGTTAAAGAG AAATGGGCTC GTCCCGAGGG TAGTGCCTGA
2451 TCAATGAGGG CGGGGCCTGG CGTCTGATCT GGGGCCGCCC TTACGGGGCA
2501 GGGCTCAGTC CTGACGCTTG CCACCTGCTC CTACCCGGCC AGGATGGCTG
2551 AGGGCGGAGT CTATTTTACG CGTCGCCCAA TGACAGGACC TGGAATGTAC
2601 TGGCTGGGGT AGGCCTCAGT GAGTCGGCCG GTCAGGGCCC GCAGCCTCGC
2651 CCCATCCACT CCGGTGCCTC CATTTAGCTG GCCAATCAGC CCAGGAGGGG
2701 CAGGTTCCCC GGGGCCGGCG CTAGGATTTG CACTAATGTT CCTCTCCCCG
2751 CGGGTGGGGG CGGGGAAATT CATATCCCCT GTTCGTCTCA TGCGCGTCCT
2801 CCGTCCCCAA TCTAAAAAGC AATTGAAAAG GTCTATGCAA TAAAGGCAGT
2851 CGCTTCATTC CTCTCAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

99045661:
Tumor necrosis factor-alpha induces adhesion molecule expression through the sphingosine kinase pathway.

98395082:
Molecular cloning and functional characterization of murine sphingosine kinase.

98241633:
Purification and characterization of rat kidney sphingosine kinase.

99178622:
Sphingosine 1-phosphate: a prototype of a new class of second messengers.

Peptide information for frame 3

```
  1 MNGHLEAEEQ QDQRPDQELT GSWGHGPRST LVRAKAMAPP PPPLAASTSL
 51 LHGEFGSYPA RGPRFALTLT SQALHIQRLR PKPEARPRGG LVPLAEVSGC
101 CTLRSRSPSD SAAYFCIYTY PRGRRGARRR ATRTFRADGA ATYEENRAEA
151 QRWATALTCL LRGLPLPGDG EITPDLLPRP PRLLLLVNPF GGRGLAWQWC
201 KNHVLPMISE AGLSFNLIQT ERQNHARELV QGLSLSEWDG IVTVSGDGLL
251 HEVLNGLLDR PDWEEAVKMP VGILPCGSGN ALAGAVNQHG GFEPALGLDL
301 LLNCSLLLCR GGGHPLDLLS VTLASGSRCF SFLSVAWGFV SDVDIQSERF
351 RALGSARFTL GTVLGLATLH TYRGRLSYLP ATVEPASPTP AHSLPRAKSE
401 LTLTPDPAPP MAHSPLHRSV SDLPLPLPQP ALASPGSPEP LPILSLNGGG
451 PELAGDWGGA GDAPLSPDPL LSSPPGSPKA ALHSPVSEGA PVIPPSSGLP
501 LPTPDARVGA STCGPPDHLL PPLGTPLPPD WVTLEGDFVL MLAISPSHLG
551 ADLVAAPHAR FDDGLVHLCW VRSGISRAAL LRLFLAMERG SHFSLGCPQL
601 GYAAARAFRL EPLTPRGVLT VDGEQVEYGP LQAQMHPGIG TLLTGPPGCP
651 GREP
```

ORF from 270 bp to 2231 bp; peptide length: 654
Category: similarity to known protein

BLASTP hits

Entry SPAC4A8_7 from database TREMBL:
gene: "SPAC4A8.07c"; product: "hypothetical protein"; S.pombe chromosome I cosmid c4A8.
Score = 301, P = 7.9e-32, identities = 68/190, positives = 109/190

Entry CEC34C6_3 from database TREMBLNEW:
product: "C34C6.5"; Caenorhabditis elegans cosmid C34C6 >TREMBL:CEC34C6_3 product: "C34C6.5"; Caenorhabditis elegans cosmid C34C6
Score = 273, P = 9.0e-29, identities = 78/265, positives = 142/265

Entry S67059 from database PIR:
hypothetical protein YOR171c - yeast (Saccharomyces cerevisiae) >TREMBL:SC55021_9 gene: "O3615"; product: "O3615p"; Saccharomyces cerevisiae cosmid pUOA1258 from chromosome 15R. >TREMBL:SCYOR170W_2 S.cerevisiae chromosome XV reading frame ORF YOR170w

Score = 253, P = 2.0e-25, identities = 70/234, positives = 116/234

Entry S51398 from database PIR:
hypothetical protein YLR260w - yeast (Saccharomyces cerevisiae)
>TREMBL:SCL8479_4 gene: "YLR260W"; product: "Ylr260wp"; Saccharomyces cerevisiae chromosome XII cosmid 8479.
Score = 251, P = 1.0e-24, identities = 62/198, positives = 103/198

Alert BLASTP hits for DKFZphfbr2_82m6, frame 3

TREMBL:AF068749_1 gene: "SPHK1b"; product: "sphingosine kinase"; Mus musculus sphingosine kinase (SPHK1b) mRNA, complete cds., N = 2, Score = 615, P = 1.2e-92

TREMBL:AF068748_1 gene: "SPHK1a"; product: "sphingosine kinase"; Mus musculus sphingosine kinase (SPHK1a) mRNA, partial cds., N = 2, Score = 616, P = 2e-92

TREMBL:ATF18E5_16 gene: "F18E5.160"; product: "putative protein"; Arabidopsis thaliana DNA chromosome 4, BAC clone F18E5 (ESSAII project), N = 2, Score = 370, P = 6.8e-33

```
>TREMBL:AF068748_1 gene: "SPHK1a"; product: "sphingosine kinase"; Mus
     musculus sphingosine kinase (SPHK1a) mRNA, partial cds.
           Length = 504

  HSPs:

 Score = 616 (92.4 bits), Expect = 2.0e-92, Sum P(2) = 2.0e-92
 Identities = 128/260 (49%), Positives = 173/260 (66%)

Query:   154 ATALTCLLRGLPLPGDGEITPDLLPRPPRLLLLVNPFGGRGLAWQWCKNHVLPMISEAGL 213
             A   C       L  + E     LLPRP R+L+L+NP GG+G A Q  ++ V P + EA +
Sbjct:   110 APVAPCQREPRDLAMEPECPRGLLPRPCRVLVLLNPQGGKGKALQLFQSRVQPFLEEAEI 169

Query:   214 SFNLIQTERQNHARELVQGLSLSEWDGIVTVSGDGLLHEVLNGLLDRPDWEEAVKMPVGI 273
             +F LI TER+NHARELV     L  WD +  +SGDGL+HEV+NGL++RPDWE A++ P+
Sbjct:   170 TFKLILTERKNHARELVCAEELGHWDALAVMSGDGLMHEVVNGLMERPDWETAIQKPLCS 229

Query:   274 LPCGSGNALAGAVNQHGGFEPALGLDLLLNCSLLLCRGGGHPLDLLSVTLASGSRCFSFL 333
             LP GSGNALA +VN + G+E      DLL+NC+LLLCR     P++LLS+  ASG R +S L
Sbjct:   230 LPGGSGNALAASVNHYAGYEQVTNEDLLINCTLLLCRRLSPMNLLSLHTASGLRLYSVL 289

Query:   334 SVAWGFVSDVDIQSERFRALGSARFTLGTVLGLATLHTYRGRLSYLPA-TVEPASPTPAH 392
             S++WGFV+DVD++SE++R LG  RFT+GT   LA+L  Y+G+L+YLP  TV  AS  PA
Sbjct:   290 SLSWGFVADVDLESEKYRRLGEIRFTVGTFFRLASLRIYQGQLAYLPVGTV--ASKRPAS 347

Query:   393 SL-PRAKSELTLTPDPAPPMAH 413
             +L  +   + L P   P  +H
Sbjct:   348 TLVQKGPVDTHLVPLEEPVPSH 369

 Score = 324 (48.6 bits), Expect = 2.0e-92, Sum P(2) = 2.0e-92
 Identities = 72/160 (45%), Positives = 100/160 (62%)

Query:   499 LPLPTPDARVGASTC---GPPDHLLPPLGTPLPPDWVTL-EGDFVLMLAISPSHLGADLV 554
             LP+ T  ++  AST    GP D  L PL  P+P   W  + E DF+L+L +  +HL ++L
Sbjct:   335 LPVGTVASKRPASTLVQKGPVDTHLVPLEEPVPSHWTVVPEQDFLLVLVLLHTHLSSELF 394

Query:   555 AAPHARFDDGLVHLCWVRSGISRAALLRLFLAMERGSHFSLGCPQLGYAAARAFRLEPLT 614
             AAP  R + G++HL +VR+G+SRAALLRLFLAM++G H  L CP L +    AFRLEP +
Sbjct:   395 AAPMGRCEAGVMHLFYVRAGVSRAALLRLFLAMQKGKHMELDCPYLVHVPVVAFRLEPRS 454

Query:   615 PRGVLTVDGEQVEYGPLQAQMHPGIGTLLTGPPGCP-GRE 653
              RGV +VDGE +    +Q Q+HP    ++ G    P GR+
Sbjct:   455 QRGVFSVDGELMVCEAVQGQVHPNYLWMVCGSRDAPSGRD 494

 Score = 37 (5.6 bits), Expect = 3.6e-62, Sum P(2) = 3.6e-62
 Identities = 8/20 (40%), Positives = 9/20 (45%)

Query:   459 GAGDAPLSPDPLLSSPPGSP 478
             G+ DAP   D     PP  P
Sbjct:   485 GSRDAPSGRDSRRGPPPEEP 504
```

Pedant information for DKFZphfbr2_82m6, frame 3

Report for DKFZphfbr2_82m6.3

```
[LENGTH]        654
[MW]            69207.45
[pI]            6.47
[HOMOL]         TREMBL:AF068749_1 gene: "SPHK1b"; product: "sphingosine kinase"; Mus musculus
sphingosine kinase (SPHK1b) mRNA, complete cds. 2e-50
[FUNCAT]        01.06.01 lipid, fatty-acid and sterol biosynthesis  [S. cerevisiae, YLR260w]
4e-20
[PROSITE]       AMIDATION        1
[PROSITE]       CAMP_PHOSPHO_SITE       1
[PROSITE]       MYRISTYL        12
[PROSITE]       CK2_PHOSPHO_SITE        6
[PROSITE]       TYR_PHOSPHO_SITE        1
[PROSITE]       GLYCOSAMINOGLYCAN       1
[PROSITE]       PKC_PHOSPHO_SITE        8
[PROSITE]       ASN_GLYCOSYLATION       1
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY    20.18 %
```

```
SEQ   MNGHLEAEEQQDQRPDQELTGSWGHGPRSTLVRAKAMAPPPPPLAASTSLLHGEFGSYPA
SEG   ................................xxxxxxxxxxxxx...............
PRD   ccchhhhhhhcccccceeeccccccccceeehhhhcccccccceeeceeeeccccccccc

SEQ   RGPRFALTLTSQALHIQRLRPKPEARPRGGLVPLAEVSGCCTLRSRSPSDSAAYFCIYTY
SEG   ............................................................
PRD   cccceeehhhhhhhhhhhcccccccccceeeeeeeceeeeeeecccccceeeeeeeeeec

SEQ   PRGRRGARRRATRTFRADGAATYEENRAEAQRWATALTCLLRGLPLPGDGEITPDLLPRP
SEG   .xxxxxxxxxxxxxxxxxxxx..................................xxxxx
PRD   ccccchhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccc

SEQ   PRLLLLVNPFGGRGLAWQWCKNHVLPMISEAGLSFNLIQTERQNHARELVQGLSLSEWDG
SEG   xxxxxx......................................................
PRD   ceeeeeeecccccchhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhccccce

SEQ   IVTVSGDGLLHEVLNGLLDRPDWEEAVKMPVGILPCGSGNALAGAVNQHGGFEPALGLDL
SEG   .......................................................xxxxx
PRD   eeeccccccceeecccccccccchhhhccceeecccccccccccccccccccccchhhhh

SEQ   LLNCSLLLCRGGGHPLDLLSVTLASGSRCFSFLSVAWGFVSDVDIQSERFRALGSARFTL
SEG   xxxxxxxxxxxxx...............................................
PRD   hhhhhccccccccccceeeeeecccccceeeeeeeeecccccceeeehhhhhhhhhhhhhc

SEQ   GTVLGLATLHTYRGRLSYLPATVEPASPTPAHSLPRAKSELTLTPDPAPPMAHSPLHRSV
SEG   ............................................................
PRD   hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   SDLPLPLPQPALASPGSPEPLPILSLNGGGPELAGDWGGAGDAPLSPDPLLSSPPGSPKA
SEG   ..xxxxxxxxxxxxxxxxx.....................xxxxxxxxxxxxxxxxxxxx
PRD   cccccccccccccccccccceeeeecccccccccccccccccccccccccccccccccce

SEQ   ALHSPVSEGAPVIPPSSGLPLPTPDARVGASTCGPPDHLLPPLGTPLPPDWVTLEGDFVL
SEG   xx.................................xxxxxxxxxxxxxxx..........
PRD   eeccccccccccccccccccccccccccccccccccccccccccccccccccccccccee

SEQ   MLAISPSHLGADLVAAPHARFDDGLVHLCWVRSGISRAALLRLFLAMERGSHFSLGCPQL
SEG   ............................................................
PRD   eeeeecccccccccccccccccccceeeeeeeccchhhhhhhhhhhhhcccceeecccch

SEQ   GYAAARAFRLEPLTPRGVLTVDGEQVEYGPLQAQMHPGIGTLLTGPPGCPGREP
SEG   ....................................xxxxxxxxxxxxxxx...
PRD   hhhhhhhhhcccccceeeecccceeeccccccccccccccceeecccccccccccc
```

Prosite for DKFZphfbr2_82m6.3

PS00001	303->307	ASN_GLYCOSYLATION	PDOC00001
PS00002	245->249	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	129->133	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	102->105	PKC_PHOSPHO_SITE	PDOC00005
PS00005	134->137	PKC_PHOSPHO_SITE	PDOC00005
PS00005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS00005	347->350	PKC_PHOSPHO_SITE	PDOC00005
PS00005	355->358	PKC_PHOSPHO_SITE	PDOC00005
PS00005	371->374	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	614->617	PKC_PHOSPHO_SITE	PDOC00005
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006

PS00006	142->146	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00006	341->345	CK2_PHOSPHO_SITE	PDOC00006
PS00006	419->423	CK2_PHOSPHO_SITE	PDOC00006
PS00007	106->115	TYR_PHOSPHO_SITE	PDOC00007
PS00008	56->62	MYRISTYL	PDOC00008
PS00008	212->218	MYRISTYL	PDOC00008
PS00008	232->238	MYRISTYL	PDOC00008
PS00008	272->278	MYRISTYL	PDOC00008
PS00008	277->283	MYRISTYL	PDOC00008
PS00008	279->285	MYRISTYL	PDOC00008
PS00008	361->367	MYRISTYL	PDOC00008
PS00008	476->482	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	574->580	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00008	640->646	MYRISTYL	PDOC00008
PS00009	122->126	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_82m6.3)

DKFZphfkd2_1j9

group: kidney derived

DKFZphfkd2_1j9.3 encodes a novel 105 amino acid protein with high similarity to Xenopus laevis XLCL2 protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes.

strong similarity to XLCL2 protein, African clawed frog

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 2955 bp
Poly A stretch at pos. 2935, polyadenylation signal at pos. 2915

```
   1 GGGGGGGGCT GAGTGCTCAG TGGAGAGCGG GGAGTTGTGT CCACCTTGCC
  51 GACGTCGCTA GCCGTGGGGC TGTCCTGGGA AGGCGGACGG CGAGCGCCCG
 101 GTGTCCGCAC TCGGCCGCCT GCCGTGCCCG TCTGCGCCCG TGTCATCCTC
 151 ACTCGGGACG CAGGGACCGT TTTTAAATCA CAGGGGCGTG TGTCAGCCTG
 201 CCCTAGGACT TCATGTCTAT ATATTTCCCC ATTCACTGCC CCGACTATCT
 251 GAGATCGGCC AAGATGACTG AGGTGATGAT GAACACCCAG CCCATGGAGG
 301 AGATCGGCCT CAGCCCCCGC AAGGATGGCC TTTCCTACCA GATCTTCCCA
 351 GACCCGTCAG ATTTTGACCG CCGCTGCAAA CTGAAGGACC GTCTGCCCTC
 401 CATAGTGGTG GAACCCACAG AAGGGGAGGT GGAGAGCGGG GAGCTCCGGT
 451 GGCCCCCTGA GGAGTTCCTG GTCCAGGAGG ATGAGCAAGA TAACTGCGAA
 501 GAGACAGCGA AAGAAAATAA AGAGCAGTAG AGTCCCTGTG GACTCCCATG
 551 GGTCATACCA GCCAGCATCT GTTCCTGAAC TGTGTTTTTC CCATCATGAC
 601 GGAAGAAGAG AGTGAGCCGC AATTGTTCTG AAAATGTCAA ACGAGGCTTC
 651 TGTTTTGCAC CTGCAGATCA CCGAGTTGGT TTTCTTTTCT TTTCTTGCCT
 701 TTTTTTTTTT TTTGAAATTT GCCGAGCAGT GGAGCCCTCT GACAATTTGC
 751 AAGGCCCTCT GAGAAAGGAA GCTGCTTAGA GCCAGGGGGT TAGTGGGTGA
 801 GGGGAGCGAG TGCTGTTTTT GAGATCATTA TCTGAACTCA GGCAGCCTAG
 851 TAGAGGCAGT GGTGGGATTC CAATGGGTCT TGGTGGGTGG GAGGTGGGGC
 901 ATGTGCAAAG CAAGCAAGGA ACATTTGGGG TAAGAAAACA AACATGAGGC
 951 AAAAGAAAAA ATACATGTTT TTAAGAAAAC ATTGAGCAGA GAACTGCAGC
1001 CAGGATGCGC TCAGCAGACA TTCACTCTGG CCGCTGGGAC ATCAGAAAAC
1051 AAAGTCTTCA TCTCTCTCTC CAGTTTCACC CACCCCACCC TTTGCTTTCA
1101 TTTCAGGTGT GTTGGTCTAT ATGACAGGGA GGAGAGTAAA GGAGAGCAGG
1151 AGCAATTGGC TGCCTGCAAA GCCAGCTGGA GGTGAAGTGC AGGAAAGGAA
1201 AGGTCACCCC ATTCTACTCC ATGGCCTCTC TGCTCCCAGC TGTGGTAGGC
1251 TCACATAGCC AGTGTGATCG GTTTTTAAGA GGCAGTGCTT TTCAGCTTTT
1301 CTCCCTGATA TATCCATTTT GCTTCCCAGC ACTTTTTAGG AGTAGTGAGA
1351 GCACTTCCTG CCCTTGTTGG AAGCCCCAGG GTGGACACTC AGCACGAAGG
1401 TCTCTCCCTT AACTGCTGCC CTTCCAAGAC TTGCTCCCGA GATGGAGTGG
1451 GCGTGGTCTT CCAGGCTGGC CCTTCCTTCT CCTCACCGCC ACCTTCCCTG
1501 CCCCAGCCCC AGCAGCCATG GGTACATGGG TCCCCAGCTC ACCTATGGAT
1551 TCCCGCCAGT CTGCCCAGCT GCAGTACTCA CGCCCCATGG GGGATCTTGG
1601 TCTGTTTTTC TTGTGGGAGC CTAGTGGAGA GCAGACGTGG CTTTTTATGT
1651 GTCTTGTTGG GGAGGTGACT TGCATGGTGG GGACAAGGCT GTCGTGGCAA
1701 CCTTGGGATC GAGTTTGAGA CTAAAGGATG TCATGAGATC CCTGGCTTCT
1751 CCCCATGTTG TTCCCGGACA AGGGCAGAAG GGAGGCATGG CAAGGGACCT
1801 CTGCTGTCCT TACTCAACAG TGGTCCTCAT CCCTCCCCAC CTCCCACTGC
1851 TTCCTGCAAG GGCACCAGTT GTATGAGAAA GTTGGCCTTT GGACTTAGGA
1901 TTTCTTATTG TAGCTAAGAG CCATCTGAAG CAGCAGGTTG CAGGACAAAT
1951 GCTTCAGTCC GCCGAGAGCA GTACCGTGTG GCCAAGAGGT GGACTCAGAG
2001 CCTTCCTTGA GCTAAACTCG GCCAACCAAG GCACGCAGCA TGTCCCCTCA
2051 GGTCTCCAGT CAGTCCAGGT TGACCCTCAG TTCTGGACGT GTGTATATAG
2101 CTGTATTTAA TACCTCAAGG TCATTGTGGC TCTGGGGATG CCAGGGCAGG
2151 AGGACGAGGG TGCGCTGTGG ACACAGCAGT CCGCGGAATT CCGTTCTGGG
2201 AAGCCAATGG TCGCCGGCAC CCCTTGCTTC CTCCCTCTGT TGTCTGCCTG
2251 TGTGACACAC ATCAATGGCA ATAACTTCTT CCAACTCCTC GCAGAAGTGG
2301 GAGAGGCCGG CAGCCTGCAC CGAGAGGGGC TTTCCTCTCT CTTGCTCCCC
2351 GCTTCGTTCT GTTTTGGCTG CAGAGAGTGG TTCATCCATA CTCTCATTCC
2401 CTCGCCTCCC CTTGTGGACG GGGGTCTTGC CTTTTCAATT CCTGTGTTTT
2451 GGTGTCTTCC CTTATCTGCT ACCCTGAATC ACCTGTCCTG GTCTTGCTGT
2501 GTGATGGGAA CATGCTTGTA AACTGCGTAA CAAATCTACT TTGTGTATGT
2551 GTCTGTTTAT GGGGGTGGTT TATTATTTTT GCTGGTCCCT AGACCACTTT
2601 GTATGACCGT TTGCAGTCTG AGCAGGCCAG GGGCTGACAG CTAATGTCAG
2651 GACCCTCAGC GGTGGAGCCT GCTGGGGGGA CCCAGCTGCT CTTGGACAAG
```

```
2701 TGGCTGAGCT CCTATCTGGC CTCCTCTTTT TTTTTTTTTT CAAGTAATTT
2751 GTGTGTATTT CTAACTGATT GTATTGAAAA AATTCCTAGT ATTTCAGTAA
2801 AAATGCCTGT TGTGAGATGA ACCTCCTGTA ACTTCTATCT GTTCTTTTTT
2851 GAGGCTCAGG GAGAAACTAG CATTTTTTTT TTTCCAAACT ACTTTTTGTC
2901 ACTGTGACAG TTGTAAATAA AGTTTGAAAA TGCTCAAAAA AAAAAAAAAA
2951 AAAAC
```

BLAST Results

Entry HSG19750 from database EMBL:
human STS A001X24.
Score = 1050, P = 1.9e-39, identities = 212/213

Entry HSG20267 from database EMBL:
human STS A005C12.
Score = 610, P = 4.1e-19, identities = 122/122

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 213 bp to 527 bp; peptide length: 105
Category: strong similarity to known protein
Classification: unset

```
  1 MSIYFPIHCP DYLRSAKMTE VMMNTQPMEE IGLSPRKDGL SYQIFPDPSD
 51 FDRRCKLKDR LPSIVVEPTE GEVESGELRW PPEEFLVQED EQDNCEETAK
101 ENKEQ
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_1j9, frame 3

PIR:S52241 XLCL2 protein - African clawed frog, N = 1, Score = 443, P = 8e-42

PIR:S52241 XLCL2 protein - African clawed frog, N = 1, Score = 443, P = 8.2e-42

>PIR:S52241 XLCL2 protein - African clawed frog
Length = 102

HSPs:

Score = 443 (66.5 bits), Expect = 8.0e-42, P = 8.0e-42
Identities = 80/104 (76%), Positives = 95/104 (91%)

```
Query:    1 MSIYFPIHCPDYLRSAKMTEVMMNTQPMEEIGLSPRKDGLSYQIFPDPSDFDRRCKLKDR 60
            MS+++PIHC DYLRSA+MTEV+MNTQ M+EIGLSPRKD  SYQIFPDPSDF+R CKLKDR
Sbjct:    1 MSVFYPIHCTDYLRSAEMTEVIMNTQSMDEIGLSPRKD--SYQIFPDPSDFERCCKLKDR 58

Query:   61 LPSIVVEPTEGEVESGELRWPPEEFLVQEDEQDNCEETAKENKE 104
            LPSIVVEPTEG+VESGELRWPPEEF+V ED++  C++T KEN++
Sbjct:   59 LPSIVVEPTEGDVESGELRWPPEEFVVDEDKEGTCDQTKKENEQ 102
```

Pedant information for DKFZphfkd2_1j9, frame 3

Report for DKFZphfkd2_1j9.3

[LENGTH]	105
[MW]	12269.78
[pI]	4.40
[HOMOL]	PIR:S52241 XLCL2 protein - African clawed frog 5e-44

[KW] Alpha_Beta

SEQ MSIYFPIHCPDYLRSAKMTEVMMNTQPMEEIGLSPRKDGLSYQIFPDPSDFDRRCKLKDR
PRD cccccccccccchhhhhhhhhhhhccccccccccccccccccceeeecccccchhhhhhhc

SEQ LPSIVVEPTEGEVESGELRWPPEEFLVQEDEQDNCEETAKENKEQ
PRD ccceeeecccccccccccccccccceeeccccchhhhhhhhhhccc

(No Prosite data available for DKFZphfkd2_1j9.3)

(No Pfam data available for DKFZphfkd2_1j9.3)

DKFZphfkd2_24a15

group: transmembrane protein

DKFZphfkd2_24a15 encodes a novel amino acid protein with similarity to C. elegans cosmid R07G3.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker for kidney cells.

similarity to C. elegans R07G3.8
membrane regions: 1
Summary DKFZphfkd2_24a15 encodes a novel 323 amino acid protein, with similarity to C. elegans R07G3.8.

similarity to C. elegans R07G3.8

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1513 bp
Poly A stretch at pos. 1494, no polyadenylation signal found

```
   1 GGGGTACTCG GCGGCGGCGG AGCGGGCGGC AGAGCAGGGC GGCGGCGACT
  51 CGCAGGGTAC CACCATCTTA AGGACAGAAA AGCTACAGGA CTCTAGGAGG
 101 CCACCGTCCT GATTTGGGAA GTCCAACTTA CTTTGGCCAG ACAGCAGCTA
 151 AGCTGGTTCA TCCCATCAGC CTGGATTGGT GAAACTGAAT CACAGGAGAT
 201 ATTTCCAGGT TTGCTGGGAT GGGAAACCTG CTCAAAGTCC TTACCAGGGA
 251 AATTGAAAAC TATCCACACT TTTTCCTGGA TTTTGAAAAT GCTCAGCCTA
 301 CAGAAGGAGA GAGAGAAATC TGGAACCAGA TCAGCGCCGT CCTTCAGGAT
 351 TCTGAGAGCA TCCTTGCAGA CCTGCAGGCT TACAAAGGCG CAGGCCCAGA
 401 GATCCGAGAT GCAATTCAAA ATCCCAATGA CATTCAGCTT CAAGAAAAAG
 451 CTTGGAATGC GGTGTGCCCT CTTGTTGTGA GGCTAAAGAG ATTTTACGAG
 501 TTTTCCATTA GACTAGAAAA AGCTCTTCAG AGTTTATTGG AATCTCTGAC
 551 TTGTCCACCC TACACACCAA CCCAACACCT GGAAAGGGAA CAGGCCCTGG
 601 CAAAGGAGTT TGCCGAAATT TTACATTTTA CCCTTCGATT CGATGAGCTG
 651 AAGATGAGGA ACCCGGCTAT TCAGAATGAC TTCAGCTACT ACAGAAGAAC
 701 AATCAGTCGC AACCGCATCA ACAACATGCA CCTAGACATT GAGAATGAAG
 751 TCAATAATGA GATGGCCAAT CGAATGTCCC TCTTCTATGC AGAAGCCACG
 801 CCAATGCTGA AAACCCTTAG CAATGCCACA ATGCACTTTG TCTCTGAAAA
 851 CAAAACTCTG CCAATAGAGA ACACCACAGA CTGCCTCAGC ACAATGACAA
 901 GTGTCTGTAA AGTCATGCTG GAAACTCCGG AGTACAGAAG TAGGTTTACG
 951 AGTGAAGAGA CCCTGATGTT CTGCATGAGG GTGATGGTGG GAGTCATCAT
1001 CCTCTATGAC CATGTCCACC CTGTGGGAGC TTTCTGCAAG ACATCCAAGA
1051 TCGATATGAA AGGCTGCATA AAAGTTTTGA AGGAGCAGGC CCCAGACAGT
1101 GTGGAGGGGC TGCTAAATGC CCTCAGGTTC ACTACAAAGC ACTTGAACGA
1151 TGAATCAACT TCCAAACAGA TTCGAGCAAT GCTTCAGTAG AGCTCTGCTC
1201 AAAGAAGAGG ATCTATGTGC TGACCTCAGA AGATGTATAT GTTTACATAA
1251 TTTAATACAG ATTGATGTTA ATACTTGTGT ATTTACATAA CCGTTTCCTT
1301 CTTGTCACTG AAATATATGG ACCTTAATTT GTATCCTGAC TGACTCAACC
1351 CAGCAGAGCA TAAATTGACT TGAGAGCCTT ACCTTTGATG TCTGAAATGA
1401 AACCCCCTTC TCCAAAGGCA AAATTCGGAG ACTTTGATCT TTGCTACTGG
1451 AGTCCTTTAA CAACATCTAT AACGATAAAA AATTCCTAAT TGTCAAAAAA
1501 AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 219 bp to 1187 bp; peptide length: 323
Category: similarity to unknown protein

```
  1 MGNLLKVLTR EIENYPHFFL DFENAQPTEG EREIWNQISA VLQDSESILA
 51 DLQAYKGAGP EIRDAIQNPN DIQLQEKAWN AVCPLVVRLK RFYEFSIRLE
101 KALQSLLESL TCPPYTPTQH LEREQALAKE FAEILHFTLR FDELKMRNPA
151 IQNDFSYYRR TISRNRINNM HLDIENEVNN EMANRMSLFY AEATPMLKTL
201 SNATMHFVSE NKTLPIENTT DCLSTMTSVC KVMLETPEYR SRFTSEETLM
251 FCMRVMVGVI ILYDHVHPVG AFCKTSKIDM KGCIKVLKEQ APDSVEGLLN
301 ALRFTTKHLN DESTSKQIRA MLQ
```

BLASTP hits

Entry CER07G3_7 from database TREMBL:
gene: "R07G3.8"; Caenorhabditis elegans cosmid R07G3.
Score = 544, P = 1.4e-52, identities = 119/323, positives = 186/323

Alert BLASTP hits for DKFZphfkd2_24a15, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_24a15, frame 3

Report for DKFZphfkd2_24a15.3

```
[LENGTH]    323
[MW]        37313.06
[pI]        5.71
[HOMOL]     TREMBL:CER07G3_7 gene: "R07G3.8"; Caenorhabditis elegans cosmid R07G3. 4e-54

[PROSITE]   MYRISTYL           1
[PROSITE]   CK2_PHOSPHO_SITE   4
[PROSITE]   TYR_PHOSPHO_SITE   1
[PROSITE]   PKC_PHOSPHO_SITE   5
[PROSITE]   ASN_GLYCOSYLATION  3
[KW]        TRANSMEMBRANE  1

SEQ  MGNLLKVLTREIENYPHFFLDFENAQPTEGEREIWNQISAVLQDSESILADLQAYKGAGP
PRD  cccccchhhhhhhccccceeeccccccccchhhhhhhhhhhhhhhcchhhhhhhhhhccccc
MEM  ............................................................

SEQ  EIRDAIQNPNDIQLQEKAWNAVCPLVVRLKRFYEFSIRLEKALQSLLESLTCPPYTPTQH
PRD  hhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccchhh
MEM  ............................................................

SEQ  LEREQALAKEFAEILHFTLRFDELKMRNPAIQNDFSYYRRTISRNRINNMHLDIENEVNN
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhcchhhhhhhhhhhhhhh
MEM  ............................................................

SEQ  EMANRMSLFYAEATPMLKTLSNATMHFVSENKTLPIENTTDCLSTMTSVCKVMLETPEYR
PRD  hhhhhhhhhhhhcchhhhhhhhhceeecccccccccccccccceeeeehhhhhhhhcccccc
MEM  ............................................................

SEQ  SRFTSEETLMFCMRVMVGVIILYDHVHPVGAFCKTSKIDMKGCIKVLKEQAPDSVEGLLN
PRD  cccccchhhhhhhhhhhheeeeeeeccccccccccccccchhhhhhhhhhcccchhhhhh
MEM  .........MMMMMMMMMMMMMMMMMMMMM..............................

SEQ  ALRFTTKHLNDESTSKQIRAMLQ
PRD  hhhhhhccccccccchhhhhhccc
MEM  .......................
```

Prosite for DKFZphfkd2_24a15.3

PS00001	202->206	ASN_GLYCOSYLATION	PDOC00001
PS00001	211->215	ASN_GLYCOSYLATION	PDOC00001
PS00001	218->222	ASN_GLYCOSYLATION	PDOC00001
PS00005	96->99	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	275->278	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005

PS00005	314->317	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00007	231->240	TYR_PHOSPHO_SITE	PDOC00007
PS00008	297->303	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfkd2_24a15.3)

DKFZphfkd2_24b15
----

group: metabolism

DKFZphfkd2_24b15 encodes a novel 612 amino acid protein with similarity to bacterial and yeast phosphoglucomutase and phosphomannomutases.

The novel protein contains a phosphoserine signature typical for phosphoglucomutase (EC 5.4.2.2) or phosphomannomutase (EC 5.4.2.8). Thus, the protein seems to be taking part in the conversion of hexose phosphates. .

The new protein can find application in modulation of hexose metabolism pathways and as a new enzyme for biotechnologic production processes.

similarity to phosphomannomutases

complete cDNA, complete cds, EST hits
potential start at bp 30 matches kozak consensus PyCNatgG,

Sequenced by GBF

Locus: map="158.8 cR from top of Chr4 linkage group"

Insert length: 2204 bp
Poly A stretch at pos. 2186, no polyadenylation signal found

```
   1 GGGCTCTGCA GCGGTAGCAC AAGCTCAGCG ATGGCGGCTC CAGAAGGCAG
  51 CGGTCTAGGC GAGGACGCCC GGCTGGACCA GGAGACCGCC CAGTGGCTGC
 101 GCTGGGACAA GAATTCCTTA ACTTTGGAGG CAGTGAAACG ACTAATAGCA
 151 GAAGGTAATA AAGAAGAACT ACGAAAATGT TTTGGGGCCC GAATGGAGTT
 201 TGGGACAGCT GGCCTCCGAG CTGCTATGGG ACCTGGAATT TCTCGTATGA
 251 ATGACTTGAC CATCATCCAG ACTACACAGG GATTTTGCAG ATACCTGGAA
 301 AAACAATTCA GTGACTTAAA GCAGAAAGGC ATCGTGATCA GTTTTGACGC
 351 CCGAGCTCAT CCATCCAGTG GGGGTAGCAG CAGAAGGTTT GCCCGACTTG
 401 CTGCAACCAC ATTTATCAGT CAGGGGATTC CTGTGTACCT CTTTTCTGAT
 451 ATAACGCCAA CCCCCTTTGT GCCCTTCACA GTATCACATT TGAAACTTTG
 501 TGCTGGAATC ATGATAACTG CATCTCACAA TCCAAAGCAG GATAATGGTT
 551 ATAAGGTCTA TTGGGATAAT GGAGCTCAGA TCATTTCTCC TCACGATAAA
 601 GGGATTTCTC AAGCTATTGA AGAAAATCTA GAACCGTGGC CTCAAGCTTG
 651 GGACGATTCT TTAATTGATA GCAGTCCACT TCTCCACAAT CCGAGTGCTT
 701 CCATCAATAA TGACTACTTT GAAGACCTTA AAAAGTACTG TTTCCACAGG
 751 AGCGTGAACA GGGAGACAAA GGTGAAGTTT GTGCACACCT CTGTCCATGG
 801 GGTGGGTCAT AGCTTTGTGC AGTCAGCTTT CAAGGCTTTT GACCTTGTTC
 851 CTCCTGAGGC TGTTCCTGAA CAGAGAGATC CGGATCCTGA GTTTCCAACA
 901 GTGAAATACC CGAATCCCGA AGAGGGGAAA GGTGTCTTGA CTTTGTCTTT
 951 TGCTTTGGCT GACAAAACCA AGGCCAGAAT TGTTTTAGCT AACGACCCGG
1001 ATGCTGATAG ACTTGCTGTG GCAGAAAAGC AAGACAGTGG TGAATGGAGG
1051 GTGTTTTCAG GCAATGAGTT GGGGGCCCTC CTGGGCTGGT GGCTTTTTAC
1101 ATCTTGGAAA GAGAAGAACC AGGATCGCAG TGCTCTCAAA GACACGTACA
1151 TGTTGTCCAG CACCGTCTCC TCCAAAATCT TGCGGGCCAT TGCCTTAAAG
1201 GAAGGTTTTC ATTTTGAGGA AACATTAACT GGCTTTAAGT GGATGGGAAA
1251 CAGAGCCAAA CAGCTAATAG ACCAGGGGAA AACTGTTTTA TTTGCATTTG
1301 AAGAAGCTAT TGGATACATG TGCTGCCCTT TTGTTCTGGA CAAAGATGGA
1351 GTCAGTGCCG CTGTCATAAG TGCAGAGTTG GCTAGCTTCC TAGCAACCAA
1401 GAATTTGTCT TTGTCTCAGC AACTAAAGGC CATTTATGTG GAGTATGGCT
1451 ACCATATTAC TAAAGCTTCC TATTTTATCT GCCATGATCA AGAAACCATT
1501 AAGAAATTAT TTGAAAACCT CAGAAACTAC GATGGAAAAA ATAATTATCC
1551 AAAAGCTTGT GGCAAATTTG AAATTTCTGC CATTAGGGAC CTTACAACTG
1601 GCTATGATGA TAGCCAACCT GATAAAAAAG CTGTTCTTCC CACTAGTAAA
1651 AGCAGCCAAA TGATCACCTT CACCTTTGCT AATGGAGGCG TGGCCACCAT
1701 GCGCACCAGT GGGACAGAGC CCAAAATCAA GTACTATGCA GAGCTGTGTG
1751 CCCCACCTGG GAACAGTGAT CCTGAGCAGC TGAAGAAGGA ACTGAATGAA
1801 CTGGTCAGTG CTATTGAAGA ACATTTTTTC CAGCCACAGA AGTACAATCT
1851 GCAGCCAAAA GCAGACTAAA ATAGTCCAGC CTTGGGTATA CTTGCATTTA
1901 CCTACAATTA AGCTGGGTTT AACTTGTTAA GCAATATTTT TAAGGGCCAA
1951 ATGATTCAAA ACATCACAGG TATTTATGTG TTTTACAAAG ACCTACATTC
2001 CTCATTGTTT CATGTTTGAC CTTTAAGGTG AAAAAAGAAA ATGGCCAAAC
2051 CCAACAAACT AACATTCCTA CTAAAAAGTT GAGCTTGGAC ATATTTTGAA
2101 TTTTTGTAAG TGAAGATTTT TAAACTGACT AACTTAAAAA AATAGATTGT
2151 AATTGATGTG CCTTAATTTG CATAAATCAT AAATGTAAAA AAAAAAAAAA
2201 AAAA
```

BLAST Results

Entry HS705145 from database EMBL:

human STS WI-6820.
Score = 1261, P = 3.6e-52, identities = 253/254

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 31 bp to 1866 bp; peptide length: 612
Category: strong similarity to known protein

```
  1 MAAPEGSGLG EDARLDQETA QWLRWDKNSL TLEAVKRLIA EGNKEELRKC
 51 FGARMEFGTA GLRAAMGPGI SRMNDLTIIQ TTQGFCRYLE KQFSDLKQKG
101 IVISFDARAH PSSGGSSRRF ARLAATTFIS QGIPVYLFSD ITPTPFVPFT
151 VSHLKLCAGI MITASHNPKQ DNGYKVYWDN GAQIISPHDK GISQAIEENL
201 EPWPQAWDDS LIDSSPLLHN PSASINNDYF EDLKKYCFHR SVNRETKVKF
251 VHTSVHGVGH SFVQSAFKAF DLVPPEAVPE QRDPDPEFPT VKYPNPEEGK
301 GVLTLSFALA DKTKARIVLA NDPDADRLAV AEKQDSGEWR VFSGNELGAL
351 LGWWLFTSWK EKNQDRSALK DTYMLSSTVS SKILRAIALK EGFHFEETLT
401 GFKWMGNRAK QLIDQGKTVL FAFEEAIGYM CCPFVLDKDG VSAAVISAEL
451 ASFLATKNLS LSQQLKAIYV EYGYHITKAS YFICHDQETI KKLFENLRNY
501 DGKNNYPKAC GKFEISAIRD LTTGYDDSQP DKKAVLPTSK SSQMITFTFA
551 NGGVATMRTS GTEPKIKYYA ELCAPPGNSD PEQLKKELNE LVSAIEEHFF
601 QPQKYNLQPK AD
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_24b15, frame 1

TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B, N = 1, Score = 1431, P = 1.6e-146

TREMBL:SPCC1840_5 gene: "SPCC1840.05c"; product: "similarity to phosphomannomutases"; S.pombe chromosome III cosmid c1840., N = 1, Score = 1210, P = 4.2e-123

PIR:S54585 hypothetical protein YMR278w - yeast (Saccharomyces cerevisiae), N = 1, Score = 1046, P = 1e-105

PIR:A71299 probable phosphomannomutase (manB) - syphilis spirochete, N = 1, Score = 697, P = 9.7e-69

>TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B
Length = 595

HSPs:

Score = 1431 (214.7 bits), Expect = 1.6e-146, P = 1.6e-146
Identities = 285/598 (47%), Positives = 393/598 (65%)

```
Query:    13 ARLDQETAQWLRWDKNSLTLEAVKRLIAEGNKEELRKCFGARMEFGTAGLRAAMGPGISR 72
             A+LD++ A WL WDKN       +++L+ E N + L+     R+ FGTAG+R+ M  G  R
Sbjct:     6 AKLDKQVADWLAWDKNDKNRNEIQKLVDEKNVDALKARMDTRLVFGTAGVRSPMQAGFGR 65

Query:    73 MNDLTIIQTTQGFCRYLEKQFSDLKQKGIVISFDARAHPSSGGSSRRFARLAATTFISQG 132
             +NDLTIIQ T GF R++    +   K  G+ I FD R +       SRRFA L+A  F+
Sbjct:    66 LNDLTIIQITHGFARHMLNVYGQPKN-GVAIGFDGRYN------SRRFAELSANVFVRNN 118

Query:   133 IPVYLFSDITPTPFVPFTVSHLKLCAGIMITASHNPKQDNGYKVYWDNGAQIISPHDKGI 192
             IPVYLFS+++PTP V +     L   AG++ITASHNPK+DNGYK YW NGAQII PHD  I
Sbjct:   119 IPVYLFSEVSPTPVVSWATIKLGCDAGLIITASHNPKEDNGYKAYWSNGAQIIGPHDTEI 178

Query:   193 SQAIEENLEPWPQAWDDSLIDSSPLLHNPSASINNDYFEDLKKYCFHRSVNRETKVKFVH 252
              +  E   +P  + WD S + SSPL H+    I+  YFE  K   F R +N  T +KF +
Sbjct:   179 VRIKEAEPQPRDEYWDLSELKSSPLFHSADVVID-PYFEVEKSLNFTREINGSTPLKFTY 237

Query:   253 TSVHGVGHSFVQSAFKAFDLVPPE--AVPEQRDPDPEFPTVKYPNPEEGKGVLTLSFALA 310
             ++ HG+G+ + +  F  F        +V EQ+DP+P+FPT+ +PNPEEG+ VLTL+   A
Sbjct:   238 SAFHGIGYHYTKRMFAEFGFPASSFISVAEQQDPNPDFPTIPFPNPEEGRKVLTLAMETA 297
```

```
Query:   311 DKTKARIVLANDPDADRLAVAEKQDSGEWRVFSGNELGALLGWWLFTSWKEKNQDRSALK 370
             DK  + ++LANDPDADR+ +AEKQ  GEWRVF+GNE+GAL+ WW++T+W++ N +  A K
Sbjct:   298 DKNGSTVILANDPDADRIQMAEKQKDGEWRVFTGNEMGALITWWIWTNWRKANPNADASK 357

Query:   371 DTYMLSSTVSSKILRAIALKEGFHFEETLTGFKWMGNRAKQLIDQGKTVLFAFEEAIGYM 430
              Y+L+S VSS+I++ IA  EGF  E TLTGFKWMGNRA++L   G  V+ A+EE+IGYM
Sbjct:   358 -VYILNSAVSSQIVKTIADAEGFKNETTLTGFKWMGNRAEELRADGNQVILAWEESIGYM 416

Query:   431 CCP-FVLDKDGVSAAVISAELASFLATKNLSLSQQLKAIYVEYGYHITKASYFICHDQET 489
              P   +DKDGVSAA + AE+A+FL  +  SL  QL A+Y  YG+H+ +++Y++    E
Sbjct:   417 --PGHTMDKDGVSAAAVFAEIAAFLHAEGKSLQDQLYALYNRYGFHLVRSTYWMVPAPEV 474

Query:   490 IKKLFENLRNYDGKNNYPKACGKFEISAIRDLTTGYDDSQPDKKAVLPTSKSSQMITFTF 549
             KKLF  LR  D K  +P   G+ E++++RDLT GYD+S+PD K VLP S SS+M+TF
Sbjct:   475 TKKLFSTLRA-DLK--FPTKIGEAEVASVRDLTIGYDNSKPDNKPVLPLSTSSEMVTFFL 531

Query:   550 ANGGVATMRTSGTEPKIKYYAELCAPPGNS--DPEQLKKELNELVSAIEEHFFQPQKYNL 607
              G V T+R SGTEPKIKYY EL   PG +  D E +  E+++L   +     +PQ++ L
Sbjct:   532 KTGSVTTLRASGTEPKIKYYIELITAPGKTQNDLESVISEMDQLEKDVVATLLRPQQFGL 591

Query:   608 QPK 610
              P+
Sbjct:   592 IPR 594
```

Pedant information for DKFZphfkd2_24b15, frame 1

Report for DKFZphfkd2_24b15.1

```
[LENGTH]     612
[MW]         68311.58
[pI]         6.28
[HOMOL]      TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B 1e-157
[FUNCAT]     01.05.01 carbohydrate utilization    [S. cerevisiae, YMR278w] 1e-111
[FUNCAT]     g carbohydrate metabolism and transport    [H. influenzae, HI0740] 3e-66
[FUNCAT]     c energy conversion    [M. genitalium, MG053] 4e-50
[FUNCAT]     m outer membrane and cell wall    [H. influenzae, HI1463] 2e-04
[BLOCKS]     BL00607D cAMP phosphodiesterases class-II proteins
[BLOCKS]     BL00710 Phosphoglucomutase and phosphomannomutase phosphoserine signa
[EC]         5.4.2.8 Phosphomannomutase 3e-56
[EC]         5.4.2.2 Phosphoglucomutase 1e-09
[PIRKW]      isomerase 3e-56
[PIRKW]      intramolecular transferase 3e-56
[SUPFAM]     Methanobacterium thermoautotrophicum phosphomannomutase 1e-06
[SUPFAM]     probable phosphorylating protein ureC 9e-06
[PROSITE]    PGM_PMM 1
[PROSITE]    MYRISTYL         10
[PROSITE]    LIPOCALIN        2
[PROSITE]    CK2_PHOSPHO_SITE       9
[PROSITE]    GLYCOSAMINOGLYCAN      1
[PROSITE]    PKC_PHOSPHO_SITE       8
[PROSITE]    ASN_GLYCOSYLATION      1
[PFAM]       Phosphoglucomutase and phosphomannomutase phosphoserine
[KW]         Alpha_Beta
```

```
SEQ    MAAPEGSGLGEDARLDQETAQWLRWDKNSLTLEAVKRLIAEGNKEELRKCFGARMEFGTA
PRD    cccccccccchhhhhhhhhhhhhhhcchhhhhhhhhhhhhhcchhhhhhhhhhhhccccc

SEQ    GLRAAMGPGISRMNDLTIIQTTQGFCRYLEKQFSDLKQKGIVISFDARAHPSSGGSSRRF
PRD    ccccccccccccccceeeeeehhhhhhhhhhhcccccceeeeeecccccccccccchhhh

SEQ    ARLAATTFISQGIPVYLFSDITPTPFVPFTVSHLKLCAGIMITASHNPKQDNGYKVYWDN
PRD    hhhhhhhhhhccceeeeecccccccccchhhhhhhcccceeeeeeccccccccceeeeecc

SEQ    GAQIISPHDKGISQAIEENLEPWPQAWDDSLIDSSPLLHNPSASINNDYFEDLKKYCFHR
PRD    ccccccccchhhhhhhhhhhhhhcccccccccccccccccccccchhhhhhhhhhhhhhcc

SEQ    SVNRETKVKFVHTSVHGVGHSFVQSAFKAFDLVPPEAVPEQRDPDPEFPTVKYPNPEEGK
PRD    ccccccceeeeeeeecccccccchhhhhhhhhcccccccccccccccccccccccccccchh

SEQ    GVLTLSFALADKTKARIVLANDPDADRLAVAEKQDSGEWRVFSGNELGALLGWWLFTSWK
PRD    hhhhhhhhhhhhhcceeeeeccccccceeeeeccccccceeeecccchhhhhhhhhhhhhhh

SEQ    EKNQDRSALKDTYMLSSTVSSKILRAIALKEGFHFEETLTGFKWMGNRAKQLIDQGKTVL
PRD    hccccccccccceeeeeeehhhhhhhhhhhcccceeeeeecccccchhhhhhhhhhccceee
```

```
SEQ    FAFEEAIGYMCCPFVLDKDGVSAAVISAELASFLATKNLSLSQQLKAIYVEYGYHITKAS
PRD    hhhhhccccccccccccccchhhhhhhhhhhhhhhhhcccchhhhhhhhhhhccccccc

SEQ    YFICHDQETIKKLFENLRNYDGKNNYPKACGKFEISAIRDLTTGYDDSQPDKKAVLPTSK
PRD    eeecccchhhhhhhhhhhhhhhccccccccchhhhhhhcccccccccccccccccccccc

SEQ    SSQMITFTFANGGVATMRTSGTEPKIKYYAELCAPPGNSDPEQLKKELNELVSAIEEHFF
PRD    ccceeeeeeccccceeeeeccccccccceeeeeecccccccchhhhhhhhhhhhhhhhhhh

SEQ    QPQKYNLQPKAD
PRD    cccccccccccc
```

Prosite for DKFZphfkd2_24b15.1

PS00001	458->462	ASN_GLYCOSYLATION	PDOC00001
PS00002	7->11	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	117->120	PKC_PHOSPHO_SITE	PDOC00005
PS00005	290->293	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00005	380->383	PKC_PHOSPHO_SITE	PDOC00005
PS00005	489->492	PKC_PHOSPHO_SITE	PDOC00005
PS00005	538->541	PKC_PHOSPHO_SITE	PDOC00005
PS00005	556->559	PKC_PHOSPHO_SITE	PDOC00005
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	343->347	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00006	523->527	CK2_PHOSPHO_SITE	PDOC00006
PS00006	528->532	CK2_PHOSPHO_SITE	PDOC00006
PS00006	560->564	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	593->597	CK2_PHOSPHO_SITE	PDOC00006
PS00008	6->12	MYRISTYL	PDOC00008
PS00008	61->67	MYRISTYL	PDOC00008
PS00008	100->106	MYRISTYL	PDOC00008
PS00008	159->165	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00008	257->263	MYRISTYL	PDOC00008
PS00008	344->350	MYRISTYL	PDOC00008
PS00008	348->354	MYRISTYL	PDOC00008
PS00008	440->446	MYRISTYL	PDOC00008
PS00008	552->558	MYRISTYL	PDOC00008
PS00710	159->174	PGM_PMM	PDOC00589
PS00213	346->358	LIPOCALIN	PDOC00187
PS00213	344->358	LIPOCALIN	PDOC00187

Pfam for DKFZphfkd2_24b15.1

```
HMM_NAME    Phosphoglucomutase and phosphomannomutase phosphoserine

HMM              *GvnVIdIGQNGMMPTPMIYFaIRTYKhmcmggGIMITaSHNPGGPDnDN
                  G+ V +      ++PTP + F +      H+++ +GIMITASHNP      DN
Query       132   GIPVYLFS--DITPTPFVPFTVS---HLKLCAGIMITASHNP--KQ-DN    172

HMM              GIK*
                 G+K
Query       173  GYK     175
```

DKFZphfkd2_24e23

group: kidney derived

DKFZphfkd2_24e23 encodes a novel 198 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes.

unknown

complete cDNA, complete cds, 1 EST hit,
many ATGs in front of the ORF

Sequenced by GBF

Locus: unknown

Insert length: 1723 bp
Poly A stretch at pos. 1695, no polyadenylation signal found

```
   1 GGGGGATTTT CGATCATGAC AACGATAGCA ATTGATATAC CTTCAAAATA
  51 CGTGTCCAGT GAGTGTTGAT TGTGTGTGGT TTCTCTAGGA GACCGTGTTC
 101 ATGCAACACA GCATTATTTC ACCGCCTTTA CCCCAGCTTC TTCATACACA
 151 TGCACTTGTC AAGGGCTCTT TGGCTGAAGA GAAGTTAGAA GTTTCCAGAT
 201 ATGGAGGGGT ATTTTCAGCA GATATGCCCA CCGCCATGGT TTTGTCAGCT
 251 CTGTAGGGTG GTCTTGCACC CTGCTCACTG CTGGCATCAC CTGAGCCTAT
 301 GGCAGATACC CAGTGCTGCC CGCCACCATG TGAATTCATC AGCTCTGCAG
 351 GCACAGACCT TGCACTAGGA ATGGGCTGGG ACGCCACCCT CTGCCTCTTA
 401 CCATTCACTG GGTTTGGCAA GTGTGCTGGG ATCTGGAATC ACATGGATGA
 451 GGAACCCGAT AATGGTGACG ACCGAGGTAG CAGGCGAACC ACTGGCCAGG
 501 GCAGGAAGTG GGCAGCTCAC GGGACTATGG CTGCACCGCG GGTTCATACC
 551 GACTACCATC CTGGAGGTGG GAGCGCATGC TCATCTGTAA AAGTCCGGTC
 601 CCACGTTGGA CACACCGGGG TCTTCTTCTT TGTTGACCAG GATCCTCTGG
 651 CAGTGTCTTT AACAAGCCAG AGTCTGATCC CACCGCTCAT AAAGCCAGGG
 701 TTGTTGAAAG CTTGGGGCTT CCTCCTCCTC TGTGCGCAGC CCTCAGCAAA
 751 CGGTCACAGC CTGTGCTGTC TGCTGTACAC CGACTTGGTA TCATCCCATG
 801 AACTGTCCCC CTTTCGTGCT CTGTGCTTAG GGCCCTCTGA TGCCCCATCT
 851 GCCTGCGCTT CCTGCAACTG TTTAGCAAGC ACCTATTATC TATAGGGTGC
 901 TGGGGTGCTG GGCGAGGCCA ATCGCTCCTA TTACTTTCTG CCCTGGGGAC
 951 GTCCTGTTTT CCCACCTACC CCTGTAACGC CTCTGCTCTG CCTTCCCATC
1001 TGCGGGCCTA ACGCCATCCC ACAAGGGCTG GGCTGTCCGT TCAGAAGAGA
1051 AACTGGGAAG GGGCCTTGAG GACCTGTGTC CAGGCAGGGT GGACAAGGGC
1101 TTTGTGCAGG GAGCTCCTCT CCCATCTTTG TGTCCTGACA GCCGTGACCG
1151 TGACCCCTCA AAGCAGAGCC AGTAGTGATC AGTATCCTGC TGCTTCAAGC
1201 CTGCACGGTC CTCTTCTCCT CTCCGCACAT CTGCATGCCT GTCAAACCCA
1251 GAGTAGTTTG GGGCCTGGTA AACAGAGGGA AGTTGGCTGG AGGAGGCCAG
1301 TCAGGAGTGC AAGAACCCCG CGTACTCTGT CCCACGTGGA TAAAGTCTCT
1351 AATTCCAGTC TGAGGTGAAT TCTTAGAGAG TGCTTTCATT TAATGTTTGC
1401 TTTATGCATT TCCCCTGCAG CTGTGACTAA TTGTGGAACA GCATACATTT
1451 TGTTTTGAGA CTCTCTTGAG ATTTTTCTGG CAGTGTAAGG TCTACACCAT
1501 TTTCCTCTCA GCATCAGAGA AGGCAGAAAG CAAGAGAAAG GAATGCAATG
1551 TGAGCAAGGC CAGGCACACT TGTGCTACTG CAGTTGGCAA GAATGGAGTC
1601 TAATCCCAGC ACTTTGGGAG GCCGAGGCGG GTGGATCACC TGAGGTCAGG
1651 AATTTGAGAC CAACCTGGCC AACATGTTGA AACCTCGTCT GTACTAAAAA
1701 TACAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 299 bp to 892 bp; peptide length: 198
Category: putative protein

```
  1 MADTQCCPPP CEFISSAGTD LALGMGWDAT LCLLPFTGFG KCAGIWNHMD
 51 EEPDNGDDRG SRRTTGQGRK WAAHGTMAAP RVHTDYHPGG GSACSSVKVR
101 SHVGHTGVFF FVDQDPLAVS LTSQSLIPPL IKPGLLKAWG FLLLCAQPSA
151 NGHSLCCLLY TDLVSSHELS PFRALCLGPS DAPSACASCN CLASTYYL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_24e23, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_24e23, frame 2

Report for DKFZphfkd2_24e23.2

```
[LENGTH]    198
[MW]        20948.98
[pI]        6.01
[PROSITE]   MYRISTYL        5
[PROSITE]   AMIDATION       1
[PROSITE]   CAMP_PHOSPHO_SITE     1
[PROSITE]   CK2_PHOSPHO_SITE      1
[PROSITE]   PKC_PHOSPHO_SITE      2
[KW]        All_Beta
[KW]        LOW_COMPLEXITY     6.06 %
```

```
SEQ   MADTQCCPPPCEFISSAGTDLALGMGWDATLCLLPFTGFGKCAGIWNHMDEEPDNGDDRG
SEG   ............................................................
PRD   ccccccccccccccccccccccccccccceeeeeccccccceeeecccccccccccccccc

SEQ   SRRTTGQGRKWAAHGTMAAPRVHTDYHPGGGSACSSVKVRSHVGHTGVFFFVDQDPLAVS
SEG   ............................................................
PRD   cccccccccccccccccccceeeeeccccccccccccceeeeeeeccccceeeeeeccccceee

SEQ   LTSQSLIPPLIKPGLLKAWGFLLLCAQPSANGHSLCCLLYTDLVSSHELSPFRALCLGPS
SEG   .....xxxxxxxxxxxx...........................................
PRD   ecccccccccccchhhhhhhhhhcccccccccceeeeeeeeecccccccccceeeecccc

SEQ   DAPSACASCNCLASTYYL
SEG   ..................
PRD   cccccccccccccccccc
```

Prosite for DKFZphfkd2_24e23.2

PS00004	62->66	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	61->64	PKC_PHOSPHO_SITE	PDOC00005
PS00005	96->99	PKC_PHOSPHO_SITE	PDOC00005
PS00006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS00008	18->24	MYRISTYL	PDOC00008
PS00008	60->66	MYRISTYL	PDOC00008
PS00008	89->95	MYRISTYL	PDOC00008
PS00008	91->97	MYRISTYL	PDOC00008
PS00008	134->140	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfkd2_24e23.2)

DKFZphfkd2_24n20
----

group: intracellular transport and trafficking

DKFZphfkd2_24n20.3 encodes a novel 366 amino acid protein with similarity to human eps8 binding protein e3B1 and spectrins.

The new protein contains an Src homology domain 3 and is similar to human eps8 SH3 domain binding protein 1 (e3B1) and spectrins. Eps8 is a substrate of receptor tyrosine kinases involved in mitogenic signaling. Spectrin is part of the submembrane cytoskeletal network in the human erythrocyte ghost. Nonerythroid spectrins are proposed to have roles in cell adhesion, establishment of cell polarity, and attachment of other cytoskeletal structures to the plasma membrane. The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton.

The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamics.

strong similarity to eps8 binding protein e3B1

complete cDNA, complete cds, few EST hits
potential start at Bp 300, but there are ATGs in other frames in
5' region of the cDNA

Sequenced by GBF

Locus: /map="17"

Insert length: 1719 bp
Poly A stretch at pos. 1699, polyadenylation signal at pos. 1680

```
   1 GGGGACAGCT GCCCCGACCT TGGCTTCCTC TGCTGGGTGG GATTGGGGGC
  51 TGGGCCCCCA AATGGGCCCC TGGCTTCCCC CTTCCTCTGG GCAGGGGACA
 101 GAGAGACACA GGCTCGGGGA GCAGGACTGA CTTCCTCTTG TCCCGGAATG
 151 AGCATGCCTG CCCTTTGCAA GCAGGTTTGG GTCTCACGCA GAGGAAACCA
 201 AAAGCAATAA GAGGGAGGGA AGGCAGAGCA ACCAATCAAG GGCAGGGTGA
 251 GACTCAAAAC GAGCGGGCTC CCTGGGGAGC CAGACAGAGG CTGGGGGTGA
 301 TGGCGGAGCT ACAGCAGCTG CAGGAGTTTG AGATCCCCAC TGGCCGGGAG
 351 GCTCTGAGGG GCAACCACAG TGCCCTGCTG CGGGTCGCTG ACTACTGCGA
 401 GGACAACTAT GTGCAGGCCA CAGACAAGCA GAAGGCGCTG GAGGAGACCA
 451 TGGCCTTCAC TACCCAGGCA CTGGCCAGCG TGGCCTACCA GGTGGGCAAC
 501 CTGGCCGGGC ACACTCTGCG CATGTTGGAC CTGCAGGGGG CCGCCCTGCG
 551 GCAGGTGGAA GCCCGTGTAA GCACGCTGGG CCAGATGGTG AACATGCATA
 601 TGGAGAAGGT GGCCCGAAGG GAGATCGGCA CCTTAGCCAC TGTCCAGCGG
 651 CTGCCCCCCG GCCAGAAGGT CATCGCCCCA GAGAACCTAC CCCCTCTCAC
 701 GCCCTACTGC AGGAGACCCC TCAACTTTGG CTGCCTGGAC GACATTGGCC
 751 ATGGGATCAA GGACCTCAGC ACGCAGCTGT CAAGAACAGG CACCCTGTCT
 801 CGAAAGAGCA TCAAGGCCCC TGCCACACCC GCCTCCGCCA CCTTGGGGAG
 851 ACCGCCCCGG ATTCCCGAGC CAGTGCACCT GCCGGTGGTG CCCGACGGCA
 901 GACTCTCCGC CGCCTCCTCT GCGTCTTCCC TGGCCTCGGC CGGCAGCGCC
 951 GAAGGTGTCG GTGGGGCCCC CACGCCCAAG GGGCAGGCAG CACCTCCAGC
1001 CCCACCTCTC CCCAGCTCCT TGGACCCACC TCCTCCACCA GCAGCCGTCG
1051 AGGTGTTCCA GCGGCCTCCC ACGCTGGAGG AGTTGTCCCC ACCCCCACCG
1101 GACGAAGAGC TGCCCCTGCC ACTGGACCTG CCTCCTCCTC CACCCCTGGA
1151 TGGAGATGAA TTGGGGCTGC CTCCACCCCC ACCAGGATTT GGGCCTGATG
1201 AGCCCAGCTG GGTGCCTGCC TCATACTTGG AGAAAGTGGT GACACTGTAC
1251 CCATACACCA GCCAGAAGGA CAATGAGCTC TCCTTCTCTG AGGGCACTGT
1301 CATCTGTGTC ACTCGCCGCT ACTCCGATGG CTGGTGCGAG GGCGTCAGCT
1351 CGGAGGGGAC TGGATTCTTC CCTGGGAACT ATGTGGAGCC CAGCTGCTGA
1401 CAGCCCAGGG CTCTCTGGGC AGCTGATGTC TGCACTGAGT GGGTTTCATG
1451 AGCCCCAAGC CAAAACCAGC TCCAGTCACA GCTGGACTGG GTCTGCCCAC
1501 CTCTTGGGCT GTGAGCTGTG TTCTGTCCTT CCTCCCATCG GAGGGAGAAG
1551 GGGTCCTGGG GAGAGAGAAT TTATCCAGAG GCCTGCTGCA GATGGGGAAG
1601 AGCTGGAAAC CAAGAAGTTT GTCAACAGAG GACCCCTACT CCATGCAGGA
1651 CAGGGTCTCC TGCTGCAAGT CCCAACTTTG AATAAAACAG ATGATGTCCA
1701 AAAAAAAAAA AAAAAAAAA
```

BLAST Results

Entry AC004797 from database EMBL:
Homo sapiens chromosome 17, clone hRPC.62_O_9, complete sequence.
Score = 2316, P = 5.9e-255, identities = 464/465
7 exons Bp 93317-110902

Medline entries

97163405:
Isolation and characterization of e3B1, an eps8 binding protein that regulates cell growth.

98256293:
Identification of a candidate human spectrin Src homology 3 domain-binding protein suggests a general mechanism of association of tyrosine kinases with the spectrin-based membrane skeleton.

Peptide information for frame 3

ORF from 300 bp to 1397 bp; peptide length: 366
Category: strong similarity to known protein

```
  1 MAELQQLQEF EIPTGREALR GNHSALLRVA DYCEDNYVQA TDKQKALEET
 51 MAFTTQALAS VAYQVGNLAG HTLRMLDLQG AALRQVEARV STLGQMVNMH
101 MEKVARREIG TLATVQRLPP GQKVIAPENL PPLTPYCRRP LNFGCLDDIG
151 HGIKDLSTQL SRTGTLSRKS IKAPATPASA TLGRPPRIPE PVHLPVVPDG
201 RLSAASSASS LASAGSAEGV GGAPTPKGQA APPAPPLPSS LDPPPPPAAV
251 EVFQRPPTLE ELSPPPPDEE LPLPLDLPPP PPLDGDELGL PPPPPGFGPD
301 EPSWVPASYL EKVVTLYPYT SQKDNELSFS EGTVICVTRR YSDGWCEGVS
351 SEGTGFFPGN YVEPSC
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_24n20, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_24n20, frame 3

Report for DKFZphfkd2_24n20.3

```
[LENGTH]      366
[MW]          38947.21
[pI]          4.93
[HOMOL]       TREMBL:U87166_1 gene: "SSH3BP1"; product: "spectrin SH3 domain binding protein
1"; Homo sapiens spectrin SH3 domain binding protein 1 (SSH3BP1) mRNA, complete cds. 3e-48

[FUNCAT]      10.99 other signal-transduction activities   [S. cerevisiae, YGR136w] 9e-06
[FUNCAT]      30.10 nuclear organization    [S. cerevisiae, YGR136w] 9e-06
[FUNCAT]      99 unclassified proteins      [S. cerevisiae, YPR154w] 3e-05
[FUNCAT]      30.04 organization of cytoskeleton   [S. cerevisiae, YDR388w] 2e-04
[FUNCAT]      03.04 budding, cell polarity and filament formation  [S. cerevisiae, YDR388w]
2e-04
[FUNCAT]      06.10 assembly of protein complexes   [S. cerevisiae, YDR162c] 4e-04
[BLOCKS]      BL50002B Src homology 3 (SH3) domain proteins profile
[SUPFAM]      SH3 homology 6e-17
[PROSITE]     MYRISTYL        6
[PROSITE]     CAMP_PHOSPHO_SITE    1
[PROSITE]     CK2_PHOSPHO_SITE     6
[PROSITE]     PKC_PHOSPHO_SITE     8
[PROSITE]     ASN_GLYCOSYLATION    1
[PFAM]        Src homology domain 3
[KW]          Irregular
[KW]          3D
[KW]          LOW_COMPLEXITY    24.04 %
```

```
SEQ    MAELQQLQEFEIPTGREALRGNHSALLRVADYCEDNYVQATDKQKALEETMAFTTQALAS
SEG    ............................................................
1aboA  ............................................................

SEQ    VAYQVGNLAGHTLRMLDLQGAALRQVEARVSTLGQMVNMHMEKVARREIGTLATVQRLPP
SEG    ............................................................
1aboA  ............................................................
```

```
SEQ     GQKVIAPENLPPLTPYCRRPLNFGCLDDIGHGIKDLSTQLSRTGTLSRKSIKAPATPASA
SEG     ............................................................
1aboA   ............................................................

SEQ     TLGRPPRIPEPVHLPVVPDGRLSAASSASSLASAGSAEGVGGAPTPKGQAAPPAPPLPSS
SEG     ......................xxxxxxxxxxxxxxxxxxxxx......xxxxxxxxxxx
1aboA   ............................................................

SEQ     LDPPPPPAAVEVFQRPPTLEELSPPPPDEELPLPLDLPPPPPLDGDELGLPPPPPGFGPD
SEG     xxxxxxxxx......xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
1aboA   ............................................................

SEQ     EPSWVPASYLEKVVTLYPYTSQKDNELSFSEGTVICVTRRYSDGWCEGVSSEGTGFFPGN
SEG     xx..........................................................
1aboA   ............EECCCBCCCTTTBCCBTTTEEEEEEEEETTTTEEEEEEETTEEEEEEGG

SEQ     YVEPSC
SEG     ......
1aboA   GEEE..
```

Prosite for DKFZphfkd2_24n20.3

PS00001	22->26	ASN_GLYCOSYLATION	PDOC00001
PS00004	339->343	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	14->17	PKC_PHOSPHO_SITE	PDOC00005
PS00005	41->44	PKC_PHOSPHO_SITE	PDOC00005
PS00005	72->75	PKC_PHOSPHO_SITE	PDOC00005
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	225->228	PKC_PHOSPHO_SITE	PDOC00005
PS00005	321->324	PKC_PHOSPHO_SITE	PDOC00005
PS00005	338->341	PKC_PHOSPHO_SITE	PDOC00005
PS00006	14->18	CK2_PHOSPHO_SITE	PDOC00006
PS00006	239->243	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	321->325	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00008	21->27	MYRISTYL	PDOC00008
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	94->100	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	215->221	MYRISTYL	PDOC00008
PS00008	332->338	MYRISTYL	PDOC00008

Pfam for DKFZphfkd2_24n20.3

```
HMM_NAME      Src homology domain 3

HMM               *pyVIALYDYqAqdpDELSFkEGDIIiIIEdsDD.WWrgRnnnTNGQEGW
                   ++V+ LY+Y++Q ++ELSF EG +I +   +  D W++G +      +G+
Query         311  EKVVTLYPYTSQKDNELSFSEGTVICVTRRYSDGWCEGVSSE---GTGF    356

HMM               IPSNYVEPi*
                  +P NYVEP
Query         357 FPGNYVEPS    365
```

DKFZphfkd2_24p5
----

group: intracellular transport and trafficking

DKFZphfkd2_24p5 encodes a novel 811 amino acid protein which is a novel splice variant of human ankyrin G.

The ankyrin 3 gene encodes a novel ankyrin, which is expressed in multiple tissues, with very high expression at the axonal initial segment and nodes of Ranvier of neurons in the central and peripheral nervous systems. Ankyrin G shows several tissue-specific alternative mRNA processing. The different ankyrin G proteins participate in maintenance/targeting of ion channels and cell adhesion molecules to nodes of Ranvier and axonal initial segments.

The new protein can find application in modulating the structure and membrane topology of Ranvier nodes and other neuronal cell membranes.

Human ankyrin G (ANK-3) new splice variant

splice variant
potential frame shift at 2720 was checked
see BLASTX

Sequenced by EMBL

Locus: /map="10q21"

Insert length: 3470 bp
Poly A stretch at pos. 3459, no polyadenylation signal found

```
   1 AGCTTTAAAA GGATGTCTGC GAAGTGGTCA AAAGGATCTT AACCTCAATT
  51 AAGTGGGGTT TTTTAAAAAG ATTTTTTGGG GGGCCTGAAA TTTTGAAAAT
 101 CTTCGAACTC TGAGTGGGGA AAGATGTATA ATTCCTCAAT TGCCTACGAG
 151 GATATCAAGA TGCTGAGAGG AATTCAGCGG TGGTGAAGAG AGTGGATACA
 201 AACCAGGGAT TGGTTTCCTT GAGCTGTTTT GGAGGTTGAT TCTAAATCAC
 251 TGCTTAAGGA ATTCCTGGAA ACATCAGGAA AACATTTGAT CATCCAAGCC
 301 TAGTGGAAAT GGCTTTACCG CAGAGTGAAG ATGCAATGAC CGGGGACACA
 351 GACAAATATC TTGGGCCACA GGACCTTAAG GAATTGGGTG ATGATTCCCT
 401 GCCTGCAGAG GGTTACATGG GCTTTAGTCT CGGAGCGCGT TCTGCCAGCC
 451 TCCGCTCCTT CAGTTCGGAT GGGTCTTACA CCTTGAACAG AAGCTCCTAT
 501 GCACGGGACA GCATGATGAT TGAAGAACTC CTCGTGCCAT CCAAAGAGCA
 551 GCATCTAACA TTCACAAGGG AATTTGATTC AGATTCTCTT AGACATTACA
 601 GCTGGGCTGC AGACACCTTA GACAATGTCA ATCTTGTTCC AAGCCCCATT
 651 CATTCTGGGT TTCTGGTTAG CTTTATGGTG GACGCGAGAG GGGGCTCCAT
 701 GAGAGGAAGC CGTCATCACG GGATGAGAAT CATCATTCCT CCACGCAAGT
 751 GTACGGCCCC CACTCGAATC ACCTGCCGTT TGGTAAAGAG ACATAAACTG
 801 GCCAACCCAC CCCCCATGGT GGAAGGAGAG GGATTAGCCA GTAGGCTGGT
 851 AGAAATGGGT CCTGCAGGGG CACAATTTTT AGGCCCTGTC ATAGTGGAAA
 901 TCCCTCACTT TGGGTCCATG AGAGGAAAAG AGAGAGAACT CATTGTTCTT
 951 CGAAGTGAAA ATGGTGAAAC TTGGAAGGAG CATCAGTTTG ACAGCAAAAA
1001 TGAAGATTTA ACCGAGTTAC TTAATGGCAT GGATGAAGAA CTTGATAGCC
1051 CAGAAGAGTT AGGGAAAAAG CGTATCTGCA GGATTATCAC GAAAGATTTC
1101 CCCCAGTATT TTGCAGTGGT TTCCCGGATT AAGCAGGAAA GCAACCAGAT
1151 TGGTCCTGAA GGTGGAATTC TGAGCAGCAC CACAGTGCCC CTTGTTCAAG
1201 CATCTTTCCC AGAGGGTGCC CTAACTAAAA GAATTCGAGT GGGCCTCCAG
1251 GCCCAGCCTG TTCCAGATGA AATTGTGAAA AAGATCCTTG GAAACAAAGC
1301 AACTTTTAGC CCAATTGTCA CTGTGGAACC AAGAAGACGG AAATTCCATA
1351 AACCAATCAC AATGACCATT CCGGTGCCCC CGCCCTCAGG AGAAGGTGTA
1401 TCCAATGGAT ACAAAGGGGA CACTACACCC AATCTGCGTC TTCTCTGTAG
1451 CATTACAGGG GGCACTTCGC CTGCTCAGTG GGAAGACATC ACAGGAACAA
1501 CTCCTTTGAC GTTTATAAAA GATTGTGTCT CCTTTACAAC CAATGTTTCA
1551 GCCAGATTTT GGCTTGCAGA CTGCCATCAA GTTTTAGAAA CTGTGGGGTT
1601 AGCCACGCAA CTGTACAGAG AATTGATATG TGTTCCATAT ATGGCCAAGT
1651 TTGTTGTTTT TGCCAAAATG AATGATCCCG TAGAATCTTC CTTGCGATGT
1701 TTCTGCATGA CAGATGACAA AGTGGACAAA ACTTTAGAGC AACAAGAGAA
1751 TTTTGAGGAA GTCGCAAGAA GCAAAGATAT TGAGGTTCTG GAAGGAAAAC
1801 CTATTTATGT TGATTGTTAT GGAAATTTGG CCCCACTTAC CAAAGGAGGA
1851 CAGCAACTTG TTTTTAACTT TTATTCTTTC AAAGAAAATA GACTGCCATT
1901 TTCCATCAAG ATTAGAGACA CCAGCCAAGA GCCCTGTGGT CGTCTGTCTT
1951 TTCTGAAAGA ACCAAAGACA ACAAAAGGAC TGCCTCAAAC AGCGGTTTGC
2001 AACTTAAATA TCACTCTGCC AGCACATAAA AAGATTGAGA AAACAGATGG
2051 ACGACAGAGC TTCGCATCCT TAGCTTTACG TAAGCGCTAC AGCTACTTGA
2101 CTGAGCCTGG AATGAGTCCA CAGAGTCCAT GTGAACGGAC AGATATCAGG
2151 ATGGCAATAG TAGCCGATCA CCTGGGACTT AGTTGGACAG AACTGGCAAG
2201 GGAACTGAAT TTTTCAGTGG ATGAAATCAA TCAAATACGT GTGGAAAATC
2251 CAAATTCTTT AATTTCTCAG AGCTTCATGT TTTTAAAAAA ATGGGTTACC
2301 AGAGACGGAA AAAATGCCAC AACTGATGCC TTAACTTCGG TCTTGACAAA
2351 AATTAATCGA ATAGATATAG TGACACTGCT AGAAGGACCA ATATTTGATT
```

```
2401 ATGGAAATAT TTCAGGCACC AGAAGTTTTG CAGATGAGAA CAATGTTTTC
2451 CATGACCCTG TTGATGGTTA TCCTTCCCTT CAAGTGGAAC TGGAAACCCC
2501 CACAGGGTTG CACTACACAC CACCTACCCC TTTCCAGCAA GATGATTATT
2551 TTAGTGATAT CTCTAGCATA GAATCTCCCC TTAGAACCCC TAGTAGACTG
2601 AGTGATGGGC TAGTGCCTTC CCAGGGGAAC ATAGAGCATT CCGCAGATGG
2651 ACCTCCAGTC GTAACTGCAG AAGACGCTTC CTTAGAAGAC AGCAAACTGG
2701 AAGACTCAGT GCCTTTAACA GAAATGCCTG AAGCAGTGAT GTAGATGAGA
2751 GCCAGTTGGA GAATGTATGT CTGAGTTGGC AGAATGAGAC ATCAAGTGGA
2801 AACCTAGAGT CCTGCGCTCA AGCTCGAAGA GTAACTGGTG GGTTACTAGA
2851 TCGACTGGAT GACAGCCCTG ACCAGTGTAG AGATTCCATT ACCTCATATC
2901 TCAAAGGAGA AGCTGGCAAA TTTGAAGCAA ATGGAAGCCA TACAGAAATC
2951 ACTCCAGAAG CAAAGACAAA ATCTTACTTT CCAGAATCCC AAAATGATGT
3001 AGGAAAAACAG AGTACCAAGG AAACTCTGAA ACCAAAAATA CATGGATCTG
3051 GTCATGTTGA AGAACCAGCA TCACCACTAG CAGCATATCA GAAATCTCTA
3101 GAAGAAACCA GCAAGCTTAT AATAGAAGAG ACTAAACCCT GTGTGCCTGT
3151 CAGTATGAAA AAGATGAGTA GGACTTCTCC AGCAGATGGC AAGCCAAGGC
3201 TTAGCCTCCA TGAAGAAGAG GGGTCCAGTG GGTCTGAGCA AAAGCAGGGA
3251 GAAGGTTTTA AGGTGAAAAC GAAGAAAGAA ATCCGGCATG TGGAAAAGAA
3301 GAGCCACTCG TAACAGCGAA CGGTCAGTCA AGGATCATAA GTTTTTACTG
3351 CCAGTATTGA GAAATTCGTG GAAGAAATGT CAGCAGGAAG TAAAAATTCA
3401 CCGAGAAGTG TGTGTGTGTT CGCTGCTTCC ACACATTAAT GGCATGATTT
3451 TTTTTATGCA AAAAAAAAAA
```

BLAST Results

Entry MMANK3A_1 from database TREMBL:
Ank3"; product: "ankyrin 3"; Mus mu... +3 4022 0.0 2

Entry HS13616 from database EMBL:
Human ankyrin G (ANK-3) mRNA, complete cds.
Length = 14,770
Plus Strand HSPs:
Score = 8505 (1276.1 bits), Expect = 0.0, Sum P(3) = 0.0
Identities = 1799/1873 (96%)

Medline entries

95394457:
Chromosomal localization of the ankyrinG gene (ANK3/Ank3) to human 10q21 and mouse 10.

95138209:
A new ankyrin gene with neural-specific isoforms localized at the axonal initial segment and node of Ranvier

Peptide information for frame 3

ORF from 309 bp to 2741 bp; peptide length: 811
Category: known protein
Classification: unset

```
  1 MALPQSEDAM TGDTDKYLGP QDLKELGDDS LPAEGYMGFS LGARSASLRS
 51 FSSDGSYTLN RSSYARDSMM IEELLVPSKE QHLTFTREFD SDSLRHYSWA
101 ADTLDNVNLV PSPIHSGFLV SFMVDARGGS MRGSRHHGMR IIIPPRKCTA
151 PTRITCRLVK RHKLANPPPM VEGEGLASRL VEMGPAGAQF LGPVIVEIPH
201 FGSMRGKERE LIVLRSENGE TWKEHQFDSK NEDLTELLNG MDEELDSPEE
251 LGKKRICRII TKDFPQYFAV VSRIKQESNQ IGPEGGILSS TTVPLVQASF
301 PEGALTKRIR VGLQAQPVPD EIVKKILGNK ATFSPIVTVE PRRRKFHKPI
351 TMTIPVPPPS GEGVSNGYKG DTTPNLRLLC SITGGTSPAQ WEDITGTTPL
401 TFIKDCVSFT TNVSARFWLA DCHQVLETVG LATQLYRELI CVPYMAKFVV
451 FAKMNDPVES SLRCFCMTDD KVDKTLEQQE NFEEVARSKD IEVLEGKPIY
501 VDCYGNLAPL TKGGQQLVFN FYSFKENRLP FSIKIRDTSQ EPCGRLSFLK
551 EPKTTKGLPQ TAVCNLNITL PAHKKIEKTD GRQSFASLAL RKRYSYLTEP
601 GMSPQSPCER TDIRMAIVAD HLGLSWTELA RELNFSVDEI NQIRVENPNS
651 LISQSFMFLK KWVTRDGKNA TTDALTSVLT KINRIDIVTL LEGPIFDYGN
701 ISGTRSFADE NNVFHDPVDG YPSLQVELET PTGLHYTPPT PFQQDDYFSD
751 ISSIESPLRT PSRLSDGLVP SQGNIEHSAD GPPVVTAEDA SLEDSKLEDS
801 VPLTEMPEAV M
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_24p5, frame 3

TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds., N = 1, Score = 4022, P = 0

TREMBL:MMANK3B_3 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (7kb isoform) mRNA, complete cds., N = 1, Score = 4005, P = 0

TREMBL:MMANK3B_4 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (7kb isoform) mRNA, complete cds., N = 1, Score = 4005, P = 0

>TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds.
Length = 1,094

HSPs:

Score = 4022 (603.5 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 769/805 (95%), Positives = 783/805 (97%)

```
Query:    1 MALPQSEDAMTGDTDKYLGPQDLKELGDDSLPAEGYMGFSLGARSASLRSFSSDGSYTLN 60
            MALP SEDA+TGDTDKYLGPQDLKELGDDSLPAEGY+GFSLGARSASLRSFSSD SYTLN
Sbjct:    1 MALPHSEDAITGDTDKYLGPQDLKELGDDSLPAEGYVGFSLGARSASLRSFSSDRSYTLN 60

Query:   61 RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLVPSPIHSGFLV 120
            RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLV SP+HSGFLV
Sbjct:   61 RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLVSSPVHSGFLV 120

Query:  121 SFMVDARGGSMRGSRHHGMRIIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL 180
            SFMVDARGGSMRGSRHHGMRIIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL
Sbjct:  121 SFMVDARGGSMRGSRHHGMRIIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL 180

Query:  181 VEMGPAGAQFLGPVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDLTELLNG 240
            VEMGPAGAQFLGPVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDL ELLNG
Sbjct:  181 VEMGPAGAQFLGPVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDLAELLNG 240

Query:  241 MDEELDSPEELGKKRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF 300
            MDEELDSPEELG KRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF
Sbjct:  241 MDEELDSPEELGTKRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF 300

Query:  301 PEGALTKRIRVGLQAQPVPDEIVKKILGNKATFSPIVTVEPRRRKFHKPITMTIPVPPPS 360
            PEGALTKRIRVGLQAQPVP+E VKKILGNKATFSPIVTVEPRRRKFHKPITMTIPVPPPS
Sbjct:  301 PEGALTKRIRVGLQAQPVPEETVKKILGNKATFSPIVTVEPRRRKFHKPITMTIPVPPPS 360

Query:  361 GEGVSNGYKGDTTPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA 420
            GEGVSNGYKGD TPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA
Sbjct:  361 GEGVSNGYKGDATPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA 420

Query:  421 DCHQVLETVGLATQLYRELICVPYMAKFVVFAKMNDPVESSLRCFCMTDDKVDKTLEQQE 480
            DCHQVLETVGLA+QLYRELICVPYMAKFVVFAK NDPVESSLRCFCMTDD+VDKTLEQQE
Sbjct:  421 DCHQVLETVGLASQLYRELICVPYMAKFVVFAKTNDPVESSLRCFCMTDDRVDKTLEQQE 480

Query:  481 NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYSFKENRLPFSIKIRDTSQ 540
            NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYSFKENRLPFSIKIRDTSQ
Sbjct:  481 NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYSFKENRLPFSIKIRDTSQ 540

Query:  541 EPCGRLSFLKEPKTTKGLPQTAVCNLNITLPAHKKIEKTDGRQSFASLALRKRYSYLTEP 600
            EPCGRLSFLKEPKTTKGLPQTAVCNLNITLPAHKK EK D RQSFASLALRKRYSYLTEP
Sbjct:  541 EPCGRLSFLKEPKTTKGLPQTAVCNLNITLPAHKKAEKADRRQSFASLALRKRYSYLTEP 600

Query:  601 GMSPQSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNSLISQSFMFLK 660
             MSPQSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNSLISQSFM LK
Sbjct:  601 SMSPQSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNSLISQSFMLLK 660

Query:  661 KWVTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENNVFHDPVDG 720
            KWVTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENNVFHDPVDG
Sbjct:  661 KWVTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENNVFHDPVDG 720

Query:  721 YPSLQVELETPTGLHYTPPTPFQQDDYFSDISSIESPLRTPSRLSDGLVPSQGNIEHSAD 780
            +PS QVELETP GL++TPP PFQQDD+FSDISSIESP RTPSRLSDGLVPSQGNIEH
Sbjct:  721 HPSFQVELETPMGLYWTPPNPFQQDDHFSDISSIESPFRTPSRLSDGLVPSQGNIEHPTG 780

Query:  781 GPPVVTAEDASLEDSKLEDSVPLTE 805
            GPPVVTAED SLEDSK++DSV +T+
```

Sbjct: 781 GPPVVTAEDTSLEDSKMDDSVTVTD 805

Pedant information for DKFZphfkd2_24p5, frame 3

Report for DKFZphfkd2_24p5.3

```
[LENGTH]   811
[MW]       90104.66
[pI]       5.40
[HOMOL]    TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial
ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds. 0.0
[BLOCKS]   BL50017B Death domain proteins profile
[PIRKW]    phosphoprotein 0.0
[PIRKW]    alternative splicing 0.0
[PIRKW]    peripheral membrane protein 0.0
[PIRKW]    cytoskeleton 0.0
[SUPFAM]   ankyrin 0.0
[SUPFAM]   ankyrin repeat homology 0.0
[SUPFAM]   unassigned ankyrin repeat proteins 0.0
[KW]       TRANSMEMBRANE  2
[KW]       LOW_COMPLEXITY    1.73 %

SEQ   MALPQSEDAMTGDTDKYLGPQDLKELGDDSLPAEGYMGFSLGARSASLRSFSSDGSYTLN
SEG   ............................................................
PRD   cccccccccccccccccccccccccccccccccccccccccccccceeeeeccccccccc
MEM   ............................................................

SEQ   RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLVPSPIHSGFLV
SEG   ............................................................
PRD   cccchhhhhhhhheeeehhhhhhhhhhhccccccccccccccccccccccccccccceee
MEM   ................................................MMMMMMMMMMMM

SEQ   SFMVDARGGSMRGSRHHGMRIIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL
SEG   ......xxxxxxxxxxxxxx........................................
PRD   eeeecccccccccccccceeeecccccccccccceeeeehhhhccccccccccccccccee
MEM   MMMMMMMMMMMMMMMM...........................................M

SEQ   VEMGPAGAQFLGPVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDLTELLNG
SEG   ............................................................
PRD   eecccccceeeceeeeeecccccccccccceeeeeecccceeeeecccccccchhhhhhc
MEM   MMMMMMMMMMMMMMMMMMMMMMMMMMMM................................

SEQ   MDEELDSPEELGKKRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF
SEG   ............................................................
PRD   cccccchhhhhhhheeeeeecccccceeeehhhhcccccccccccccccceeeeeeeeccc
MEM   ............................................................

SEQ   PEGALTKRIRVGLQAQPVPDEIVKKILGNKATFSPIVTVEPRRRKFHKPITMTIPVPPPS
SEG   ............................................................
PRD   ccchhhhhhhhhhhhhccccceeeecccccccccccceeeccccccccccceeeeccccc
MEM   ............................................................

SEQ   GEGVSNGYKGDTTPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA
SEG   ............................................................
PRD   cccccccccccccccceeeeeeeeecccccccccccccccceeeeeeccccccccceeeec
MEM   ............................................................

SEQ   DCHQVLETVGLATQLYRELICVPYMAKFVVFAKMNDPVESSLRCFCMTDDKVDKTLEQQE
SEG   ............................................................
PRD   cchhhhhhhhhhhhhhhhhhhcchhhhheeecccchhhhhhhhcccchhhhhhhhhhhhc
MEM   ............................................................

SEQ   NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYSFKENRLPFSIKIRDTSQ
SEG   ............................................................
PRD   cceeeccccеeeeeecccеeeeeccccccchhhhhhhhhchhhhhhhcceeeeeeecccc
MEM   ............................................................

SEQ   EPCGRLSFLKEPKTTKGLPQTAVCNLNITLPAHKKIEKTDGRQSFASLALRKRYSYLTEP
SEG   ............................................................
PRD   ccccceeeecccccccccccccccccccccccccccccccccchhhhhhhhhhhhheeecc
MEM   ............................................................

SEQ   GMSPQSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNSLISQSFMFLK
SEG   ............................................................
PRD   ccccccchhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhcceeeeeccccchhhhhhhhhh
MEM   ............................................................
```

```
SEQ   KWVTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENNVFHDPVDG
SEG   ............................................................
PRD   hhhhcccccccchhhhhhhhhcceeeeeeecccccccccccccccccccccccccccccc
MEM   ............................................................

SEQ   YPSLQVELETPTGLHYTPPTPFQQDDYFSDISSIESPLRTPSRLSDGLVPSQGNIEHSAD
SEG   ............................................................
PRD   cccceeeeccccccccccccccccccccceeeccccccccccccccccccccccccccccc
MEM   ............................................................

SEQ   GPPVVTAEDASLEDSKLEDSVPLTEMPEAVM
SEG   ...............................
PRD   ccceeeeccccccccccccccccccccccccc
MEM   ...............................
```

(No Prosite data available for DKFZphfkd2_24p5.3)

(No Pfam data available for DKFZphfkd2_24p5.3)

DKFZphfkd2_3i13
----

group: transmembrane protein

DKFZphfkd2_3i13 encodes a novel 406 amino acid protein with C. elegans cosmid Y37D8A and A. thaliana H71412 hypothetical protein.

The novel protein contains 3 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker for kidney cells.

similarity to A.thaliana and C.elegans;
membrane regions: 3

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="17"

Insert length: 2052 bp
Poly A stretch at pos. 2032, no polyadenylation signal found

```
   1 AGTGACGTGA GCGGGTTCCG GTTGTCTGGA GCCCAGCGGC GGGTGTGAGA
  51 GTCCGTAAGG AGCAGCTTCC AGGATCCTGA GATCCGGAGC AGCCGGGGTC
 101 GGAGCGGCTC CTCAAGAGTT ACTGATCTAT GAAATGGCAG AGAATGGAAA
 151 AAATTGTGAC CAGAGACGTG TAGCAATGAA CAAGGAACAT CATAATGGAA
 201 ATTTCACAGA CCCCTCTTCA GTGAATGAAA AGAAGAGGAG GGAGCGGGAA
 251 GAAAGGCAGA ATATTGTCCT GTGGAGACAG CCGCTCATTA CCTTGCAGTA
 301 TTTTTCTCTG GAAATCCTTG TAATCTTGAA GGAATGGACC TCAAAATTAT
 351 GGCATCGTCA AAGCATTGTG GTGTCTTTTT TACTGCTGCT TGCTGTGCTT
 401 ATAGCTACGT ATTATGTTGA AGGAGTGCAT CAACAGTATG TGCAACGTAT
 451 AGAGAAACAG TTTCTTTTGT ATGCCTACTG GATAGGCTTA GGAATTTTGT
 501 CTTCTGTTGG GCTTGGAACA GGGCTGCACA CCTTTCTGCT TTATCTGGGT
 551 CCACATATAG CCTCAGTTAC ATTAGCTGCT TATGAATGCA ATTCAGTTAA
 601 TTTTCCCGAA CCACCCTATC CTGATCAGAT TATTTGTCCA GATGAAGAGG
 651 GCACTGAAGG AACCATTTTT TTGTGGAGTA TCATCTCAAA AGTTAGGATT
 701 GAAGCCTGCA TGTGGGGTAT CGGTACAGCA ATCGGAGAGC TGCCTCCATA
 751 TTTCATGGCC AGAGCAGCTC GCCTCTCAGG TGCTGAACCA GATGATGAAG
 801 AGTATCAGGA ATTTGAAGAG ATGCTGGAAC ATGCAGAGTC TGCACAAGAC
 851 TTTGCCTCCC GGGCCAAACT GGCAGTTCAA AAACTAGTAC AGAAAGTTGG
 901 ATTTTTTGGA ATTTTGGCCT GTGCTTCAAT TCCAAATCCT TTATTTGATC
 951 TGGCTGGAAT AACGTGTGGA CACTTTCTGG TACCTTTTTG GACCTTCTTT
1001 GGTGCAACCC TAATTGGAAA AGCAATAATA AAAATGCATA TCCAGAAAAT
1051 TTTTGTTATA ATAACATTCA GCAAGCACAT AGTGGAGCAA ATGGTGGCTT
1101 TCATTGGTGC TGTCCCCGGC ATAGGTCCAT CTCTGCAGAA GCCATTTCAG
1151 GAGTACCTGG AGGCTCAACG GCAGAAGCTT CACCACAAAA GCGAAATGGG
1201 CACACCACAG GGAGAAAACT GGTTGTCCTG GATGTTTGAA AAGTTGGTCG
1251 TTGTCATGGT GTGTTACTTC ATCCTATCTA TCATTAACTC CATGGCACAA
1301 AGTTATGCCA AACGAATCCA GCAGCGGTTG AACTCAGAGG AGAAAACTAA
1351 ATAAGTAGAG AAAGTTTTAA ACTGCAGAAA TTGGAGTGGA TGGGTTCTGC
1401 CTTAAATTGG GAGGACTCCA AGCCGGGAAG GAAAATTCCC TTTTCCAACC
1451 TGTATCAATT TTTACAACTT TTTTCCTGAA AGCAGTTTAG TCCATACTTT
1501 GCACTGACAT ACTTTTTCCT TCTGTGCTAA GGTAAGGTAT CCACCCTCGA
1551 TGCAATCCAC CTTGTGTTTT CTTAGGGTGG AATGTGATGT TCAGCAGCAA
1601 ACTTGCAACA GACTGGCCTT CTGTTTGTTA CTTTCAAAAG GCCCACATGA
1651 TACAATTAGA GAATTCCCAC CGCACAAAAA AAGTTCCTAA GTATGTTAAA
1701 TATGTCAAGC TTTTTAGGCT TGTCACAAAT GATTGCTTTG TTTTCCTAAG
1751 TCATCAAAAT GTATATAAAT TATCTAGATT GGATAACAGT CTTGCATGTT
1801 TATCATGTTA CAATTTAATA TTCCATCCTG CCCAACCCTT CCTCTCCCAT
1851 CCTCAAAAAA GGGCCATTTT ATGATGCATT GCACACCCTC TGGGGAAATT
1901 GATCTTTAAA TTTTGAGACA GTATAAGGAA AATCTGGTTG GTGTCTTACA
1951 AGTGAGCTGA CACCATTTTT TATTCTGTGT ATTTAGGATG AAGTCTTGAA
2001 AAAAACTTTA TAAAGACATC TTTAATCATT CCAAAAAAAA AAAAAAAAAA
2051 AA
```

BLAST Results

Entry AC004686 from database EMBL:
*** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 17, clone hRPC.1073_F_15; HTGS phase 1, 8 unordered pieces.
Score = 4142, P = 6.1e-199, identities = 830/832

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 1351 bp; peptide length: 406
Category: similarity to unknown protein

```
  1 MAENGKNCDQ RRVAMNKEHH NGNFTDPSSV NEKKRREREE RQNIVLWRQP
 51 LITLQYFSLE ILVILKEWTS KLWHRQSIVV SFLLLLAVLI ATYYVEGVHQ
101 QYVQRIEKQF LLYAYWIGLG ILSSVGLGTG LHTFLLYLGP HIASVTLAAY
151 ECNSVNFPEP PYPDQIICPD EEGTEGTIFL WSIISKVRIE ACMWGIGTAI
201 GELPPYFMAR AARLSGAEPD DEEYQEFEEM LEHAESAQDF ASRAKLAVQK
251 LVQKVGFFGI LACASIPNPL FDLAGITCGH FLVPFWTFFG ATLIGKAIIK
301 MHIQKIFVII TFSKHIVEQM VAFIGAVPGI GPSLQKPFQE YLEAQRQKLH
351 HKSEMGTPQG ENWLSWMFEK LVVVMVCYFI LSIINSMAQS YAKRIQQRLN
401 SEEKTK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_3i13, frame 2

TREMBL:CEY37D8A_20 gene: "Y37D8A.22"; Caenorhabditis elegans cosmid Y37D8A, N = 1, Score = 905, P = 8.8e-91

TREMBL:ATAC98_2 gene: "YUP8H12.2"; Arabidopsis thaliana chromosome 1 YAC yUP8H12 complete sequence., N = 1, Score = 470, P = 1.1e-44

PIR:H71412 hypothetical protein - Arabidopsis thaliana, N = 1, Score = 293, P = 6e-24

```
>TREMBL:CEY37D8A_20 gene: "Y37D8A.22"; Caenorhabditis elegans cosmid
     Y37D8A
          Length = 457

  HSPs:

 Score = 905 (135.8 bits), Expect = 8.8e-91, P = 8.8e-91
 Identities = 167/317 (52%), Positives = 228/317 (71%)

Query:    38 REERQNIVLWRQPLITLQYFSLEILVILKEWTSKLWHRQSIVVSFLLLLAVLIATYYVEG 97
             R ER+ IV WR+P I + Y +EI + E K+ +++++ + + + + Y+ G
Sbjct:    93 RMERETIVFWRRPHIVIPYALMEIAHLAVELFFKILAHKTVLLLTAISIGLAVYGYHAPG 152

Query:    98 VHQQYVQRIEKQFLLYAYWIGLGILSSVGLGTGLHTFLLYLGPHIASVTLAAYECNSVNF 157
              HQ++VQ IEK L +++W+ LG+LSS+GLG+GLHTFL+YLGPHIA+VT+AAYEC S++F
Sbjct:   153 AHQEHVQTIEKHILWWSWWVLLGVLSSIGLGSGLHTFLIYLGPHIAAVTMAAYECQSLDF 212

Query:   158 PEPPYPDQIICPDEEGTEGTIFLWSIISKVRIEACMWGIGTAIGELPPYFMARAARLSGA 217
             P+PPYP+ I CP + + F W I++KVR+E+ +WG GTA+GELPPYFMARAAR+SG
Sbjct:   213 PQPPYPESIQCPSTKSSIAVTF-WQIVAKVRVESLLWGAGTALGELPPYFMARAARISGQ 271

Query:   218 EPDDEEYQEFEEMLE-HAESAQD----FASRAKLAVQKLVQKVGFFGILACASIPNPLFD 272
             EPDDEEY+EF E++ ES D RAK V+ + ++GF GIL ASIPNPLFD
Sbjct:   272 EPDDEEYREFLELMNADKESDADQKLSIVERAKSWVEHNIHRLGFPGILLFASIPNPLFD 331

Query:   273 LAGITCGHFLVPFWTFFGATLIGKAIIKMHIQKIFVIITFSKHIVEQMVAFIGAVPGIGP 332
             LAGITCGHFLVPFW+FFGATLIGKA++KMH+Q FVI+ FS H E V + +P +GP
Sbjct:   332 LAGITCGHFLVPFWSFFGATLIGKALVKMHVQMGFVILAFSDHHAENFVKILEKIPAVGP 391

Query:   333 SLQKPFQEYLEAQRQKLH 350
             +++P + LE QR+ LH
Sbjct:   392 YIRQPISDLLEKQRKALH 409
```

Pedant information for DKFZphfkd2_3i13, frame 2

Report for DKFZphfkd2_3i13.2

```
[LENGTH]        406
[MW]            46298.17
[pI]            6.47
[HOMOL]         TREMBL:CEY37D8A_20 gene: "Y37D8A.22"; Caenorhabditis elegans cosmid Y37D8A le-
79
[PROSITE]       MYRISTYL        10
[PROSITE]       CK2_PHOSPHO_SITE        3
[PROSITE]       PKC_PHOSPHO_SITE        1
[PROSITE]       ASN_GLYCOSYLATION       1
[KW]            TRANSMEMBRANE  3
[KW]            LOW_COMPLEXITY      9.85 %

SEQ   MAENGKNCDQRRVAMNKEHHNGNFTDPSSVNEKKRREREERQNIVLWRQPLITLQYFSLE
SEG   ..............................xxxxxxxxxx....................
PRD   cccccchhhhhhhhhhhcccccccccccchhhhhhhhhhhhhhhhccccchhhhhhhh
MEM   ...............................MMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ   ILVILKEWTSKLWHRQSIVVSFLLLLAVLIATYYVEGVHQQYVQRIEKQFLLYAYWIGLG
SEG   ........................................................xxxx
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeecchhhhhhhhhhhhhhhhhhh
MEM   MM....................MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM........

SEQ   ILSSVGLGTGLHTFLLYLGPHIASVTLAAYECNSVNFPEPPYPDQIICPDEEGTEGTIFL
SEG   xxxxxxxxxxx.................................................
PRD   hcccccccccceeeeeeeecchhhhhhhhhhhccccccccccccccccccccccccceeee
MEM   ............................................................

SEQ   WSIISKVRIEACMWGIGTAIGELPPYFMARAARLSGAEPDDEEYQEFEEMLEHAESAQDF
SEG   ...................................xxxxxxxxxxxxxxx........
PRD   eehhhhhhhhhhhhhcccccccccccchhhhhhhhccccchhhhhhhhhhhhhhhhhhhh
MEM   ............................................................

SEQ   ASRAKLAVQKLVQKVGFFGILACASIPNPLFDLAGITCGHFLVPFWTFFGATLIGKAIIK
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhcceeeeeeeecccccccccccccccceeeeeeehhhhhhhhhhh
MEM   ..........MMMMMMMMMMMMMMMMMMMMM.............................

SEQ   MHIQKIFVIITFSKHIVEQMVAFIGAVPGIGPSLQKPFQEYLEAQRQKLHHKSEMGTPQG
SEG   ............................................................
PRD   hhhhheeeeeeechhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhcccccccc
MEM   ............................................................

SEQ   ENWLSWMFEKLVVVMVCYFILSIINSMAQSYAKRIQQRLNSEEKTK
SEG   ..............................................
PRD   cchhhhhhhhhheeehhhhhhhhhhhhhhhhhhhhhhhhhhhcccc
MEM   ..............................................
```

Prosite for DKFZphfkd2_3i13.2

PS00001	23->27	ASN_GLYCOSYLATION	PDOC00001
PS00005	69->72	PKC_PHOSPHO_SITE	PDOC00005
PS00006	29->33	CK2_PHOSPHO_SITE	PDOC00006
PS00006	215->219	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00008	120->126	MYRISTYL	PDOC00008
PS00008	126->132	MYRISTYL	PDOC00008
PS00008	173->179	MYRISTYL	PDOC00008
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	197->203	MYRISTYL	PDOC00008
PS00008	259->265	MYRISTYL	PDOC00008
PS00008	275->281	MYRISTYL	PDOC00008
PS00008	325->331	MYRISTYL	PDOC00008
PS00008	329->335	MYRISTYL	PDOC00008
PS00008	356->362	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfkd2_3i13.2)

DKFZphfkd2_3o17
----

group: metabolism

DKFZphfkd2_3o17 encodes a novel 72 amino acid protein with similarity to bos taurus NADH-ubiquinone oxidoreductase B33 subunit (EC 1.6.5.3) (EC 1.6.99.3).

NADH:ubiquinone oxidoreductase is the first enzyme in the respiratory electron transport chain of mitochondria. It is a a membrane-bound multi-subunit protein. The bovine heart enzyme contains about 40 different polypeptides. The novel protein is the human orthologue of bovine B22.

The new protein can find application in modulation of the respiratory electron transport chain pathways of mitochondria.

strong similarity to bovine NADH-UBIQUINONE OXIDOREDUCTASE B22 subunit

complete cDNA, complete cds, EST hits,
in frame stop codon at ~274 will be checked
ESTs HS1291620/AA883920 show no stop codon at this side

Sequenced by BMFZ

Locus: unknown

Insert length: 693 bp
Poly A stretch at pos. 670, polyadenylation signal at pos. 659

```
  1 CAGCAGGCGT GCAGTTTCCC GGCTCTCCGC GCGGCCGGGG AAGGTCAGCG
 51 CCGTAATGGC GTTCTTGGCG TCGGGACCCT ACCTGACCCA TCAGCAAAAG
101 GTGTTGCGGC TTTATAAGCG GGCGCTACGC CACCTCGAGT CGTGGTGCGT
151 CCAGAGAGAC AAATACCGAT ACTTTGCTTG TTTGATGAGA GCCCGGTTTG
201 AAGAACATAA GAATGAAAAG GATATGGCGA AGGCCACCCA GCTGCTGAAG
251 GAGGCCGAGG AAGAATTCTG GTAACGTCAG CATCCACAGC CATACATCTT
301 CCCTGACTCT CCTGGGGGCA CCTCCTATGA GAGATACGAT TGCTACAAGG
351 TCCCAGAATG GTGCTTAGAT GACTGGCATC CTTCTGAGAA GGCAATGTAT
401 CCTGATTACT TTGCCAAGAG AGAACAGTGG AAGAAACTGC GGAGGGAAAG
451 CTGGGAACGA GAGGTTAAGC AGCTGCAGGA GGAAACGCCA CCTGGTGGTC
501 CTTTAACTGA AGCTTTGCCC CCTGCCCGAA AGGAAGGTGA TTTGCCCCCA
551 CTGTGGTGGT ATATTGTGAC CAGACCCCGG GAGCGGCCCA TGTAGAAAGA
601 GAGAGACCTC ATCTTTCATG CTTGCAAGTG AAATATGTTA CAGAACATGC
651 ACTTGCCCTA ATAAAAAATC AGTAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

Entry S28256 from database PIR:
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain CI-B22 - bovine
>TREMBL:MIBTCIB22_1 gene: "cI-B22"; product: "NADH-ubiquinone oxidoreductase complex B22 subunit"; B.taurus mitochondrion cI-B22 mRNA for B22 subunit of the NADH-ubiquinone oxidoreductase complex
Score = 933, P = 5.2e-93, identities = 163/179, positives = 172/179, frame +2

Medline entries

92389317
Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from RT bovine heart mitochondria. Application of a novel strategy for RT sequencing proteins using the polymerase chain reaction

Peptide information for frame 2

ORF from 56 bp to 271 bp; peptide length: 72
Category: strong similarity to known protein

```
  1 MAFLASGPYL THQQKVLRLY KRALRHLESW CVQRDKYRYF ACLMRARFEE
 51 HKNEKDMAKA TQLLKEAEEE FW*RQHPQPY IFPDSPGGTS YERYDCYKVP
101 EWCLDDWHPS EKAMYPDYFA KREQWKKLRR ESWEREVKQL QEETPPGGPL
151 TEALPPARKE GDLPPLWWYI VTRPRERPM
```

BLASTP hits

```
Sequences producing significant alignments:                       (bits)  Value

 sp|Q02369|NI2M_BOVIN|0D36CE17281FB735 (NDUFB9..)NADH-UBIQUINONE...  141  7e-34
 tr|U41534|Q18036|D34BCCB6E8FBCD5F (C16A3.4)SIMILAR TO NADH-UBIQ...   53  3e-07

 >sp|Q02369|NI2M_BOVIN|0D36CE17281FB735 (NDUFB9..)NADH-UBIQUINONE
           OXIDOREDUCTASE B22 SUBUNIT (EC 1.6.5.3) (EC 1.6.99.3)
           (COMPLEX I-B22) (CI-B22).[BOS TAURUS]
           Length = 178

  Score = 141 bits (351), Expect = 7e-34
  Identities = 63/71 (88%), Positives = 68/71 (95%)

 Query: 2  AFLASGPYLTHQQKVLRLYKRALRHLESWCVQRDKYRYFACLMRARFEEHKNEKDMAKAT 61
            AFL+SG YLTHQQKVLRLYKRALRHLESWC+ RDKYRYFACL+RARF+EHKNEKDM KAT
 Sbjct: 1  AFLSSGAYLTHQQKVLRLYKRALRHLESWCIHRDKYRYFACLLRARFDEHKNEKDMVKAT 60

 Query: 62 QLLKEAEEEFW 72
            QLL+EAEEEFW
 Sbjct: 61 QLLREAEEEFW 71


 >tr|U41534|Q18036|D34BCCB6E8FBCD5F (C16A3.4)SIMILAR TO
           NADH-UBIQUINONE OXIDOREDUCTASE B22.[CAENORHABDITIS
           ELEGANS]
           Length = 163

  Score = 52.7 bits (124), Expect = 3e-07
  Identities = 25/64 (39%), Positives = 41/64 (64%), Gaps = 1/64 (1%)

 Query: 10 LTHQQKVLRLYKRALRHLESWCVQRD-KYRYFACLMRARFEEHKNEKDMAKATQLLKEAE 68
            L+H+QKV RLYKR LR +++W    + + R+  C++RARF+ + +E D  K+  LL +
 Sbjct: 12 LSHRQKVTRLYKRCLREVDNWYGGNNLEVRFQKCIIRARFDANADEVDTRKSQILLADGC 71

 Query: 69 EEFW 72
             + W
 Sbjct: 72 RQLW 75
```

Alert BLASTP hits for DKFZphfkd2_3o17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_3o17, frame 2

Report for DKFZphfkd2_3o17.2

```
[LENGTH]     72
[MW]         8839.28
[pI]         9.26
[HOMOL]      PIR:S28256 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain CI-B22 - bovine
2e-34
[KW]         All_Alpha

SEQ    MAFLASGPYLTHQQKVLRLYKRALRHLESWCVQRDKYRYFACLMRARFEEHKNEKDMAKA
PRD    ccccccccchhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhcchhhhhhh

SEQ    TQLLKEAEEEFW
PRD    hhhhhhhhhccc
```

(No Prosite data available for DKFZphfkd2_3o17.2)

(No Pfam data available for DKFZphfkd2_3o17.2)

DKFZphfkd2_46a6
----

group: kidney derived

DKFZphfkd2_46a6 encodes a novel 315 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="228.6 cR from top of Chr15 linkage group"

Insert length: 2774 bp
Poly A stretch at pos. 2751, polyadenylation signal at pos. 2732

```
   1 CTCGCGAGCG CAGCTATGGC TGCTGGCGTA CCCTGTGCGT TAGTCACCAG
  51 CTGCTCCTCC GTCTTCTCAG GAGACCAGCT GGTCCAACAT ACCCTTGGAA
 101 CAGAAGATCT TATTGTGGAA GTGACTTCCA ATGATGCTGT GAGATTTTAT
 151 CCCTGGACCA TTGATAATAA ATACTATTCA GCAGACATCA ATCTATGTGT
 201 GGTGCCAAAC AAATTTCTTG TTACTGCAGA GATTGCAGAA TCTGTCCAAG
 251 CATTTGTGGT TTACTTTGAC AGCACACGAA AATCGGGCCT TGATAGTGTC
 301 TCCTCATGGC TTCCACTGGC AAAAGCATGG TTACCTGAGG TGATGATCTT
 351 GGTCTGCGAT AGAGTGTCTG AAGATGGTAT AAACCGACAA AAAGCTCAAG
 401 AATGGAGCCT CAAACATGGC TTTGAATTGG TAGAACTTAG TCCAGAGGAG
 451 TTGCCTGAGG AGGATGATGA CTTCCCAGAA TCTACAGGAG TAAAGCGAAT
 501 TGTCCAAGCC CTGAATGCCA ATGTGTGGTC CAATGTAGTG ATGAAGAATG
 551 ATAGGAACCA AGGCTTTAGC CTTCTCAACT CATTGACTGG AACAAACCAT
 601 AGCATTGGGT CAGCAGATCC CTGTCACCCA GAGCAACCCC ATTTGCCAGC
 651 AGCAGATAGT ACTGAATCCC TCTCTGATCA TCGGGGTGGT GCATCTAACA
 701 CAACAGATGC CCAGGTTGAT AGCATTGTGG ATCCCATGTT AGATCTGGAT
 751 ATTCAAGAAT TAGCCAGTCT TACCACTGGA GGAGGAGATG TGGAGAATTT
 801 TGAAAGACCC TTTTCAAAGT TAAAGGAAAT GAAAGACAAG GCTGCGACGC
 851 TTCCTCATGA GCAAAGAAAA GTGCATGCAG AAAAGGTGGC CAAAGCATTC
 901 TGGATGGCAA TCGGGGGAGA CAGAGATGAA ATTGAAGGCC TTTCATCTGA
 951 TGGAGAGCAC TGAATTATTC ATACTAGGGT TTGACCAACA AAGATGCTAG
1001 CTGTCTCTGA GATACCTCTC TACTCAGCCC AGTCATATTT TGCCAAAATT
1051 GCCCTTATCA TGTTGGCTGC CTGACTTGTT TATAGGGTCC CCTTAATTTT
1101 AGTTTTTAGT AGGAGGTTAA GGAGAAATCT TTTTTTTCCT CAGTATATTG
1151 TAAGAGAGTG AGGAATACAG TGATAGTAAT GAGTGAGGAT TTCTTAAATA
1201 TACTTTTTTT TTGTTCTAGG AATGAGGGTA GGATAAATCT CAGAGGTCTG
1251 TGTGATTTAC TCAAGTTGAA GACAACCTCC AGGCCATTCC TGGTCAACCT
1301 TTTAAGTAGC ATTTCCAGCA TTCACACTTG ATACTGCACA TCAGGAGTTG
1351 TGTCACCTTT CCTGGGTGAT TTGGGTTTTC TCCATTCAAG GAGCTTGTAG
1401 CTCTGAGCTA TGATGCTTTT ATTGGGAGGA AAGGAGGCAG CTGCAGAATT
1451 GATGTGAGCT ATGTGGGGCC GAAGTCTCAG CCCGCAGCTA AGTCTCTACC
1501 TAAGAAAATG CCTCTGGGCA TTCTTTTGAA GTATAGTGTC TGAGCTCATG
1551 CTAGAAAGAA TCAAAAAGCC AGTGTGGATT TTTAGGCTGT AATAAATGAG
1601 GCAAAGGATT TCTATTCCAG TGGGAAGGAA ACCTCTCTAC TGAGTTGTGG
1651 GGGATATGTT GTATGTTAGA GAGAACCTTA AGGAGTCCTT GTATGGGCCA
1701 TGGAGACAGT ATGTGATAAC ATACCGTGAT TTTCATGAAG AAATTCTTCT
1751 GTCCTAGAGT TCTCCCCTGC TGCTTGAGAT GCCAGAGCTG TGTTGTTGCA
1801 CACCTGCAAA ACAAGGCACA TTTCCCCCTT TCTCTTTAAA GCCAAAGAGA
1851 GATCACTGCC AAAGTGGGAG CACTAAGGGG TGGGTGGGGA AGTGAAATGT
1901 TAGGCGATGA ATTCCTGAGC ACCTTGTTTT TCTTCCAAGG TTCGTAGCTC
1951 CTCTCTGCCC TTCCAAGCCT GTAACCTCGG AGGACTATCT TTTGTTCTCT
2001 ATCCTTTGTC TTGTTAGAGT GGGTCAGCCC CAGAGGAACT GATAAGCAAA
2051 TGGCAAGTTT TTAAAGGAAG AGTGGAAAGT ACTGCAAATA AAAATCCTTA
2101 TTTGTTTTTG TAGACTTTGT AATGCATATC ATTAGCCCTC ACTGTGATCA
2151 TTACTGCTGT GGCTCTGAAC TGGCACATAG TACAGTGGAT GGAAGGTGCC
2201 CGCACACCAG CTGAGAACTG GTTCTGGCCT AGGTGGGCTC TAGAACCATT
2251 TACACAGCAT GAAAGAAACA GGTTGGGTTA GGAGCAGAAA GAAATAAGGC
2301 TCACACCCCT CCAGACACTA CCTTATAAGC ACTGCAGAAC CTGAAACAGA
2351 TGGCAGAAGG AATGGAATGC TACAGGGGCC AGCAGGAGTG ACCACAGGGA
2401 GGGGACAGCT CAGTGACTGG AGCATTCAGG AAGAGGCTTT CCAGGGAACA
2451 CTGGACATTG CTTAGTGACC TTTTGTTCCT TTTTTTTTTT TTTTCTTTTA
2501 CTGTTCTGAA AGACTTTGAG TCTGTGGTTC ACCACCAGCC CATCAGTGTT
2551 TCTTTGAGGT GATTGCATTA GGGAAGTTGG CTCTGGGATT GCAAAAAAAA
2601 AAAAAAGGTG GAACATGTTT TCCTTAAAAG ATGGAAGGTT TTAGAAAATA
2651 TACTAGGCCA TCTGGTTAGA AAAAACAGAC CAGACTAGAA AAAGCTGTGA
```

```
2701 ATTTGATTTT GTAGATTAAA CAAAGCCAGA TGATTAAAAT GTGATTTATT
2751 TATAAAAAAA AAAAAAAAAA AAAA
```

BLAST Results

Entry HS463358 from database EMBL:
human STS WI-14364.
Length = 472
Minus Strand HSPs:
Score = 1605 (240.8 bits), Expect = 5.0e-68, P = 5.0e-68
Identities = 347/361 (96%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 16 bp to 960 bp; peptide length: 315
Category: putative protein
Classification: unset

```
  1 MAAGVPCALV TSCSSVFSGD QLVQHTLGTE DLIVEVTSND AVRFYPWTID
 51 NKYYSADINL CVVPNKFLVT AEIAESVQAF VVYFDSTRKS GLDSVSSWLP
101 LAKAWLPEVM ILVCDRVSED GINRQKAQEW SLKHGFELVE LSPEELPEED
151 DDFPESTGVK RIVQALNANV WSNVVMKNDR NQGFSLLNSL TGTNHSIGSA
201 DPCHPEQPHL PAADSTESLS DHRGGASNTT DAQVDSIVDP MLDLDIQELA
251 SLTTGGGDVE NFERPFSKLK EMKDKAATLP HEQRKVHAEK VAKAFWMAIG
301 GDRDEIEGLS SDGEH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_46a6, frame 1

PIR:T04362 probable GTP-binding protein yptm3 - maize, N = 1, Score = 87, P = 0.21

PIR:S71585 GTP-binding protein GB2 - Arabidopsis thaliana, N = 1, Score = 86, P = 0.27

>PIR:T04362 probable GTP-binding protein yptm3 - maize
Length = 210

HSPs:

Score = 87 (13.1 bits), Expect = 2.4e-01, P = 2.1e-01
Identities = 34/160 (21%), Positives = 67/160 (41%)

```
Query:   48 TIDNKYYSADINLCVVPNKFL-VTAEIAESVQAFVVYFDSTRKSGLDSVSSWLPLAKAWL 106
            TIDNK     I        F  +T          ++ +D TR+  + ++SWL  A+
Sbjct:   49 TIDNKPIKLQIWDTAGQESFRSITRSYYRGAAGALLVYDITRRETFNHLASWLEDARQHA 108

Query:  107 PE---VMIL--VCDRVSEDGINRQKAQEWSLKHGFELVELSPEELPEEDDDFPESTGVKR 161
               VM++   CD      ++ ++ ++++ +HG   +E S +     ++ F ++ G
Sbjct:  109 NANMTVMLIGNKCDLSHRRAVSYEEGEQFAKEHGLVFMEASAKTAQNVEEAFIKTAGT-- 166

Query:  162 IVQALNANVWSNVVMKNDRNQGFSLLNSLTGTNHSIGSADPC 203
            I + +    ++        N     G+++ NS  G    S    A  C
Sbjct:  167 IYKKIQDGIFDVSNESNGIKVGYAVPNSSGGGAGSSSQAGGC 208
```

Pedant information for DKFZphfkd2_46a6, frame 1

Report for DKFZphfkd2_46a6.1

[LENGTH] 315

```
[MW]        34505.54
[pI]        4.55
[KW]        Alpha_Beta
[KW]        LOW_COMPLEXITY      6.67 %

SEQ    MAAGVPCALVTSCSSVFSGDQLVQHTLGTEDLIVEVTSNDAVRFYPWTIDNKYYSADINL
SEG    ............................................................
PRD    ccccccceeeecccccccccceeeecccccceeeeeeccccceeeecccccccccccccee

SEQ    CVVPNKFLVTAEIAESVQAFVVYFDSTRKSGLDSVSSWLPLAKAWLPEVMILVCDRVSED
SEG    ............................................................
PRD    eeeccccchhhhhhhhhhheeeeeeeccccccccccccccccccccccceeeecccccccc

SEQ    GINRQKAQEWSLKHGFELVELSPEELPEEDDDFPESTGVKRIVQALNANVWSNVVMKNDR
SEG    ...............xxxxxxxxxxxxxxxxxxxxx........................
PRD    cchhhhhhhhhhccccceeeecccccccccccccccccccchhhhhhhhccccceeeeeeccc

SEQ    NQGFSLLNSLTGTNHSIGSADPCHPEQPHLPAADSTESLSDHRGGASNTTDAQVDSIVDP
SEG    ............................................................
PRD    cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

SEQ    MLDLDIQELASLTTGGGDVENFERPFSKLKEMKDKAATLPHEQRKVHAEKVAKAFWMAIG
SEG    ............................................................
PRD    hhhhhhhhhhhcccccccccccchhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhc

SEQ    GDRDEIEGLSSDGEH
SEG    ...............
PRD    ccccccccccccccc
```

(No Prosite data available for DKFZphfkd2_46a6.1)

(No Pfam data available for DKFZphfkd2_46a6.1)

DKFZphfkd2_46b10
----

group: kidney derived

DKFZphfkd2_46b10.1 encodes a novel 315 amino acid protein with similarity to C.elegans cosmide F25B5.3

The novel protein contains a HTH-LYSR-family PROSITE pattern. Proteins of the lysR family are bacterial transcriptional regulatory proteins which bind DNA using a helix-turn-helix motif. Most of these proteins are transcription activators and usually negatively regulate their own expression. They all possess a potential 'helix-turn-helix' DNA-binding motif in their N-terminal section. The 'helix-turn-helix' motif is missing in DKFZphfkd2_46a6.1.
No informative BLAST results, no predictive PFAM or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to C.elegans F25B5.3

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 1285 bp
Poly A stretch at pos. 1266, no polyadenylation signal found

```
   1 CAGTCTACGC GAGCTGCCTG TTTTTTTCCT GCTTGGACGC GCATGAGGGC
  51 CCCGTCCATG GACCGCGCGG CCGTGGCGAG GGTGGGCGCG GTAGCGAGCG
 101 CCAGCGTGTG CGCCCTGGTG GCGGGGGTGG TGCTGGCTCA GTACATATTC
 151 ACCTTGAAGA GGAAGACGGG GCGGAAGACC AAGATCATCG AGATGATGCC
 201 AGAATTCCAG AAAAGTTCAG TTCGAATCAA GAACCCTACA AGAGTAGAAG
 251 AAATTATCTG TGGTCTTATC AAAGGAGGAG CTGCCAAACT TCAGATAATA
 301 ACGGACTTTG ATATGACACT CAGTAGATTT TCATATAAAG GGAAAAGATG
 351 CCCAACATGT CATAATATCA TTGACAACTG TAAGCTGGTT ACGGATGAAT
 401 GTAGAAAAAA GTTATTGCAA CTAAAGGAAA AATATTACGC TATTGAAGTT
 451 GATCCTGTTC TTACTGTAGA AGAGAAGTAC CCTTATATGG TGGAATGGTA
 501 TACTAAATCA CATGGTTTGC TTGTTCAGCA AGCTTTACCA AAAGCTAAAC
 551 TTAAAGAAAT TGTGGCAGAA TCTGACGTTA TGCTCAAAGA AGGATATGAG
 601 AATTTCTTTG ATAAGCTCCA ACAACATAGC ATCCCCGTGT TCATATTTTC
 651 GGCTGGAATC GGCGATGTAC TAGAGGAAGT TATTCGTCAA GCTGGTGTTT
 701 ATCATCCCAA TGTCAAAGTT GTGTCCAATT TTATGGATTT TGATGAAACT
 751 GGGGTGCTCA AAGGATTTAA AGGAGAACTA ATTCATGTAT TTAACAAACA
 801 TGATGGTGCC TTGAGGAATA CAGAATATTT CAATCAACTA AAAGACAATA
 851 GTAACATAAT TCTTCTGGGA GACTCCCAAG GAGACTTAAG AATGGCAGAT
 901 GGAGTGGCCA ATGTTGAGCA CATTCTGAAA ATTGGATATC TAAATGATAG
 951 AGTGGATGAG CTTTTAGAAA AGTACATGGA CTCTTATGAT ATTGTTTTAG
1001 TACAAGATGA ATCATTAGAA GTAGCCAACT CTATTTTACA GAAGATTCTA
1051 TAAACAAGCA TTCTCCAAGA AGACCTCTCT CCTGTGGGTG CAATTGAACT
1101 GTTCATCCGT TCATCTTGCT GAGAGACTTA TTTATAATAT ATCCTTACTC
1151 TCGAAGTGTT CCCTTTGTAT AACTGAAGTA TTTTCAGATA TGGTGAATGC
1201 ATTGACTGGA AGCTCCTTTT CTCCACCTCT CTCAACACAC TCCTCACCGT
1251 ATCTTTTAAC CCATTTAAAA AAAAAAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 43 bp to 1050 bp; peptide length: 336
Category: similarity to unknown protein
Classification: unset
Prosite motifs: HTH_LYSR_FAMILY (16-47)

```
  1 MRAPSMDRAA VARVGAVASA SVCALVAGVV LAQYIFTLKR KTGRKTKIIE
 51 MMPEFQKSSV RIKNPTRVEE IICGLIKGGA AKLQIITDFD MTLSRFSYKG
101 KRCPTCHNII DNCKLVTDEC RKKLLQLKEK YYAIEVDPVL TVEEKYPYMV
151 EWYTKSHGLL VQQALPKAKL KEIVAESDVM LKEGYENFFD KLQQHSIPVF
201 IFSAGIGDVL EEVIRQAGVY HPNVKVVSNF MDFDETGVLK GFKGELIHVF
251 NKHDGALRNT EYFNQLKDNS NIILLGDSQG DLRMADGVAN VEHILKIGYL
301 NDRVDELLEK YMDSYDIVLV QDESLEVANS ILQKIL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_46b10, frame 1

SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III., N = 1, Score = 524, P = 2.2e-50

TREMBL:AC005499_12 gene: "T6A23.12"; Arabidopsis thaliana chromosome II BAC T6A23 genomic sequence, complete sequence., N = 2, Score = 194, P = 1.4e-26

>SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III.
Length = 376

HSPs:

```
Score = 524 (78.6 bits), Expect = 2.2e-50, P = 2.2e-50
Identities = 112/300 (37%), Positives = 174/300 (58%)

Query:  44 RKTKIIEMMPEFQ--KSSVRIKNPTRVEEIICGLIKGGAAKLQIITDFDMTLSRFSYK-G 100
           +KT ++ ++  +   +  + + +PT V   +  ++ GGA K  +I+DFD TLSRF+ + G
Sbjct:  73 KKTDVVPLLMNYLLGEEQILVADPTAVAAKLRKMVVGGAGKTVVISDFDYTLSRFANEQG 132

Query: 101 KRCPTCHNIID-NCKLVTDECRKKLLQLKEKYYAIEVDPVLTVEEKYPYMVEWYTKSHGL 159
           +R  T H + D N   +  E  +K + LK KYY IE  P LT+EEK P+M +W+  SH L
Sbjct: 133 ERLSTTHGVFDDNVMRLKPELGQKFVDLKNKYYPIEFSPNLTMEEKIPHMEKWWGTSHSL 192

Query: 160 LVQQALPKAKLKEIVAESDVMLKEGYENFFDKLQQHSIPVFIFSAGIGDVLEEVIRQA-G 218
           +V +    K  +++ V +S ++ K+G E+F + L  H+IP+ IFSAGIG+++E  ++Q  G
Sbjct: 193 IVNEKFSKNTIEDFVRQSRIVFKDGAEDFIEALDAHNIPLVIFSAGIGNIIEYFLQQKLG 252

Query: 219 VYHPNVKVVSNFMDFDETGVLKGFKGELIHVFNKHDGAL-RNTEYFNQLKDNSNIILLGD 277
               N  +SN + FDE      F    LIH F K+  + + T +F+ +    N+ILLGD
Sbjct: 253 AIPRNTHFISNMILFDEDDNACAFSEPLIHTFCKNSSVIQKETSFFHDIAGRVNVILLGD 312

Query: 278 SQGDLRMADGVANVEHILKIGYLNDRVDEL--LEKYMDSYDIVLVQDESLEVANSILQKI 335
           S GD+ M  GV     LK+GY N +D+  L+ Y + YDIVL+ D +L VA  I+ I
Sbjct: 313 SMGDIHMDVGVERDGPTLKVGYYNGSLDDTAALQHYEEVYDIVLIHDPTLNVAQKIVDII 372
```

Pedant information for DKFZphfkd2_46b10, frame 1

Report for DKFZphfkd2_46b10.1

```
[LENGTH]     336
[MW]         37948.37
[pI]         6.67
[HOMOL]      SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III.
3e-51
[PROSITE]    HTH_LYSR_FAMILY       1
[KW]         TRANSMEMBRANE  2
[KW]         LOW_COMPLEXITY       7.44 %

SEQ   MRAPSMDRAAVARVGAVASASVCALVAGVVLAQYIFTLKRKTGRKTKIIEMMPEFQKSSV
SEG   .......xxxxxxxxxxxxxxxxxxxxxxxxx............................
PRD   cccchhhhhcchhhhhhheeehhhhhhhhhhhhhhhhhhhhhhcccceeeehhhhhhhhhee
MEM   .............MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.................

SEQ   RIKNPTRVEEIICGLIKGGAAKLQIITDFDMTLSRFSYKGKRCPTCHNIIDNCKLVTDEC
SEG   ............................................................
PRD   eecccchhhhhhhhhccccceeeeecccccceeeecccccccccccccccccchhhhhh
MEM   ............................................................
```

```
SEQ    RKKLLQLKEKYYAIEVDPVLTVEEKYPYMVEWYTKSHGLLVQQALPKAKLKEIVAESDVM
SEG    ............................................................
PRD    hhhhhhhhhhheeeccccccccccccchhhhhhccccchhhhhhccchhhhhhhhhhhcc
MEM    ............................................................

SEQ    LKEGYENFFDKLQQHSIPVFIFSAGIGDVLEEVIRQAGVYHPNVKVVSNFMDFDETGVLK
SEG    ............................................................
PRD    ccccchhhhhhhhhcccceeeeecccchhhhhhhhhhcccccceeeeeeeccccccccccee
MEM    .............MMMMMMMMMMMMMMMMMM.............................

SEQ    GFKGELIHVFNKHDGALRNTEYFNQLKDNSNIILLGDSQGDLRMADGVANVEHILKIGYL
SEG    ............................................................
PRD    ecceeeeeeeecccccccccccchhhhhhhceeeeeccccccccccccccccceeeeeec
MEM    ............................................................

SEQ    NDRVDELLEKYMDSYDIVLVQDESLEVANSILQKIL
SEG    ....................................
PRD    cchhhhhhhhhhhhheeeeeecchhhhhhhhhhhccc
MEM    ....................................
```

Prosite for DKFZphfkd2_46b10.1

PS00044	16->47	HTH_LYSR_FAMILY	PDOC00043

(No Pfam data available for DKFZphfkd2_46b10.1)

DKFZphfkd2_46d13
----

group: kidney derived

DKFZphfkd2_46d13 encodes a novel 506 amino acid protein with weak similarity to KE03 protein

The novel protein contains a RGD site.
No informative BLAST results; No predictive prosite, pfam or SCOP motive

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to KE03 protein

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="227.6 cR from top of Chr1 linkage group"

Insert length: 3346 bp
Poly A stretch at pos. 3328, polyadenylation signal at pos. 3308

```
   1 CTCTCGCGAG AGGAGCAAGA GGAAGATGGC CGTGCCCTGT TTTTCGGTGT
  51 AAGGCAGCAG ACGGCGGCTG CGACGGCGAG ACTGAGATCC TGGTGTCGTG
 101 GGCACCTGAG TTCTAGCTTC CCCCAGCGAG CGCGCGTCCC TTCGTGCCTA
 151 GGCGAGAGCC GGCTCTTCCC CGGGAGATGC GTTTGTCCCA GGCTCGGGGG
 201 CTCAGTGGGA GTTCATGCTG CGCTGGAGGC TCTTGGCCAC CGCTCTAATC
 251 GCCTTGTGCC GCCGCAGCGC CAGCTCCGTC GCCAGCGGTG AGCCTCCCGA
 301 TTCCCCCCCT TGCCCCTGGC GGCGGCGATG ACCGGGGAGA AGATCCGCTC
 351 ACTGCGGAGG GACCACAAGC CCAGCAAAGA AGAAGGGGAC CTGCTGGAGC
 401 CCGGGGATGA AGAAGCGGCG GCTGCCCTCG GCGGTACCTT TACCAGAAGC
 451 AGGATTGGCA AGGGCGGCAA AGCTTGTCAT AAGATCTTCA GTAACCATCA
 501 CCACCGGCTA CAGCTGAAGG CAGCTCCGGC CTCCTCCAAT CCCCCCGGCG
 551 CCCCGGCTCT GCCGCTGCAC AATTCCTCCG TGACTGCCAA CTCCCAGTCC
 601 CCGGCCCTTC TGGCCGGCAC CAACCCCGTT GCTGTCGTCG CGGATGGAGG
 651 CAGTTGCCCC GCACACTACC CGGTGCACGA GTGCGTCTTC AAGGGGGATG
 701 TGAGGAGACT CTCCTCTCTC ATCCGCACGC ACAATATCGG GCAGAAAGAT
 751 AATCACGGAA ATACTCCTTT ACACCTTGCT GTGATGTTAG GAAATAAAGT
 801 TACAGCTCTT TTGAGGAAGC TTAAGCAGCA ATCCAGGGAA AGTGTTGAAG
 851 AAAAACGACC TCGATTATTA AAAGCCCTGA AAGAGCTAGG TGACTTTTAT
 901 CTAGAACTTC ACTGGGATTT TCAAAGCTGG GTGCCTTTAC TTTCCCGAAT
 951 TCTGCCTTCC GATGCATGTA AAATATACAA ACAAGGTATC AATATCAGGC
1001 TTGACACAAC TCTCATAGAC TTTACTGACA TGAAGTGCCA ACGAGGGGAT
1051 CTAAGCTTCA TTTTCAATGG GGATGCGGCG CCCTCTGAAT CTTTTGTAGT
1101 ATTAGACAAT GAACAAAAAG TTTATCAGCG AATACATCAT GAGGAATCAG
1151 AGATGGAAAC AGAAGAAGAG GTGGATATTT TAATGAGCAG TGATATTTAC
1201 TCTGCAACTT TATCAACAAA ATCAATTTCT TTCACGCGTG CCCAGACAGG
1251 ATGGCTTTTT CGGGAAGATA AAACAGAAAG AGTAGGAAAC TTTTTGGCAG
1301 ACTTTTACCT GGTGAATGGA CTTGTTATAG AATCAAGGAA AAGAAGAGAA
1351 CATCTCAGTG AAGAGGATAT TCTTCGAAAT AAGGCCATCA TGGAGAGTTT
1401 GAGTAAAGGT GGAAACATAA TGGAACAGAA TTTTGAGCCG ATTCGAAGAC
1451 AGTCTCTTAC ACCGCCTCCT CAGAACACTA TTACATGGGA AGAATATATA
1501 TCTGCTGAAA ATGGAAAAGC TCCTCATCTG GGTAGAGAAT TGGTGTGCAA
1551 AGAGAGTAAG AAAACGTTTA AAGCTACGAT AGCCATGAGC CAGGAATTTC
1601 CCTTAGGGAT AGAGTTATTA TTGAATGTTT TAGAAGTAGT AGCTCCCTTC
1651 AAGCACTTTA ACAAGCTTAG AGAATTTGTT CAGATGAAGC TTCCTCCAGG
1701 CTTTCCTGTA AAATTAGATA TACCTGTGTT TCCCACAATC ACAGCCACTG
1751 TGACTTTTCA GGAGTTTCGA TACGATGAAT TTGATGGCTC CATCTTTACT
1801 ATACCTGATG ACTACAAGGA AGACCCAAGC CGTTTTCCTG ATCTTTAACT
1851 GACGTGGAAA AGGATGCCGT CTAACCAAGG AAAGAAAATA CAGAGACCCT
1901 AGAAGTGGAT CCAAATAGAA GGGACAAATG CTTTCAGTGA AGAAAAGGGA
1951 ATTACACATT GAATCGACAC ATCAGTAATA CGATACAGTG AAATGGGCCT
2001 CTAATAAGAA TTTCAGCGAG TTTTCTGATG TGCCATTTTT TGTCTTTTTA
2051 AAAATATACA TATTATAAAT GTAATAGTTT GACACATTAA TGACCCTAAG
2101 ACCTGCGTAT GTGAAGCAGC TATGAGTGCT GTGATTTGTT TTTAAAAATT
2151 TTTACACTTC TTGTTGAAAT ATATATGCAT ATAAATATAT CTATATCTAT
2201 ATCTATATCT AAAACACTCC TGGACCATTA ACGTAAATTA AATGTCTTAA
2251 GAGATATGGA GCCCTTTTAA ACTTGTCATC TTTATGCAAG GTGACATTTA
2301 TAAATATTCC TTCGAGCTTT GTTTTCATAA AATGTAAACT ATGTAACATT
2351 ATGTATAGTT CAGTAATTTG AATGTTTGTT CAATATAATG AACTAGAAGG
2401 AATGCAATTT TCTGTAGATG AATGAACCAA ATGGTAACCA TTAAACAATT
2451 GCATTTATAT GTTGCAATAC ATTTCAGAAG GAGCGTTCAC TCTGCAGGGA
2501 ATAAGGTACC TCCTTTAGCA CCTTAGTGCA ATTCATTGTG GTGCTATTTG
2551 TTTTTACCTG AATGTTTGTT ACTAATCTTC CTTTCATAGA ACCTCTATTT
2601 TTTTTTTTTC TAAACTTGAG TTTGAGTCCT TGTTATGGTC ATCATAAGGT
```

```
2651 AATGGTTAGC ATGTTTAAAG ATATTCCTCT TCCAAATCTC AGCACTTTAA
2701 AAAAAAATCC AAATTTTTAA ACTTGCTTCC TAATAAGTAC ACATCGGTCT
2751 GATTATTTTG TTTGTTTTTA GTAGAATATG GATGCATTGG TGTCAGTTTT
2801 AAAAAACAAT ACACATATTT TGGACAACCC TACATATTTA ATCCTTTCAA
2851 AATAAGATAA AAACATTTTA TATGCTAACA GAATATATTT GTTACAAGTT
2901 AAAGTCCAGA AGTATACACA AGATTGATTA CTCCTATTAT TTTTTTTAAA
2951 TCACAGGAAA ATATTGATTT CATTGTCTCC AAAGTGATAA AATCTTGTAT
3001 TACTCATTTT TGCACTTAAA ATTTTTCTTA TTTATTCCAA GGTGGTTTGA
3051 AGGTCCAAGT ATGAAAATAA ATTAGGGGGA TTAATGTATA ACAGTTATAA
3101 AGTATCATGT TGTATTAAAG AGCTTACTTA GATTGATGTT TTTAAAATGT
3151 ATCCTGATGA ATGTCTCAAG AATGCATCTG TCAAGTTTTT TAGACTGACC
3201 AGTAGCTTAA ACTTTTTTCA GGATTTTAGG TAATTTGAAA GGAGTTTAGA
3251 GACCCTTATT GAAAATATGA TTTAAAAATC CAAAGCATAA ACCGTAAGAA
3301 AAATTTTAAA TAAACATCTT TAAAGCTGAA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HS121353 from database EMBL:
human STS WI-14729.
Score = 1697, P = 1.9e-69, identities = 363/379

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 328 bp to 1845 bp; peptide length: 506
Category: similarity to unknown protein

```
  1 MTGEKIRSLR RDHKPSKEEG DLLEPGDEEA AAALGGTFTR SRIGKGGKAC
 51 HKIFSNHHHR LQLKAAPASS NPPGAPALPL HNSSVTANSQ SPALLAGTNP
101 VAVVADGGSC PAHYPVHECV FKGDVRRLSS LIRTHNIGQK DNHGNTPLHL
151 AVMLGNKVTA LLRKLKQQSR ESVEEKRPRL LKALKELGDF YLELHWDFQS
201 WVPLLSRILP SDACKIYKQG INIRLDTTLI DFTDMKCQRG DLSFIFNGDA
251 APSESFVVLD NEQKVYQRIH HEESEMETEE EVDILMSSDI YSATLSTKSI
301 SFTRAQTGWL FREDKTERVG NFLADFYLVN GLVIESRKRR EHLSEEDILR
351 NKAIMESLSK GGNIMEQNFE PIRRQSLTPP PQNTITWEEY ISAENGKAPH
401 LGRELVCKES KKTFKATIAM SQEFPLGIEL LLNVLEVVAP FKHFNKLREF
451 VQMKLPPGFP VKLDIPVFPT ITATVTFQEF RYDEFDGSIF TIPDDYKEDP
501 SRFPDL
```

BLASTP hits

Entry CEC01F1_3 from database TREMBL:
gene: "C01F1.6"; Caenorhabditis elegans cosmid C01F1.
Score = 371, P = 4.5e-61, identities = 69/138, positives = 96/138

Entry CEC18F10_9 from database TREMBL:
gene: "C18F10.7"; Caenorhabditis elegans cosmid C18F10.
Score = 383, P = 3.4e-39, identities = 103/349, positives = 182/349

Entry AF064604_1 from database TREMBL:
product: "KE03 protein"; Homo sapiens KE03 protein mRNA, partial cds.
Score = 348, P = 8.3e-32, identities = 95/295, positives = 148/295

Alert BLASTP hits for DKFZphfkd2_46d13, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_46d13, frame 1

Report for DKFZphfkd2_46d13.1

[LENGTH]	506
[MW]	57003.12
[pI]	6.40

[HOMOL] TREMBL:CEC18F10_9 gene: "C18F10.7"; Caenorhabditis elegans cosmid C18F10. 2e-35

[BLOCKS]	BL01288E	
[PROSITE]	RGD	1
[PROSITE]	MYRISTYL	7
[PROSITE]	CAMP_PHOSPHO_SITE	2
[PROSITE]	CK2_PHOSPHO_SITE	9
[PROSITE]	PKC_PHOSPHO_SITE	6
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	Alpha_Beta	
[KW]	LOW_COMPLEXITY	7.51 %

```
SEQ  MTGEKIRSLRRDHKPSKEEGDLLEPGDEEAAAALGGTFTRSRIGKGGKACHKIFSNHHHR
SEG  .................xxxxxxxxxxxx...............................
PRD  ccceeeeecccccccccccccccccchhhhhhcccccccccccccceeeeeeecchhhh

SEQ  LQLKAAPASSNPPGAPALPLHNSSVTANSQSPALLAGTNPVAVVADGGSCPAHYPVHECV
SEG  ....xxxxxxxxxxxxxxxx........................................
PRD  hhhhhhccccccccceeeccccccccccccccceeeccccccceeeecccccccccccceee

SEQ  FKGDVRRLSSLIRTHNIGQKDNHGNTPLHLAVMLGNKVTALLRKLKQQSRESVEEKRPRL
SEG  ............................................................
PRD  ecchhhhhhhhhhcccccccccccccceeeecccchhhhhhhhhhhcchhhhhhhhhh

SEQ  LKALKELGDFYLELHWDFQSWVPLLSRILPSDACKIYKQGINIRLDTTLIDFTDMKCQRG
SEG  ............................................................
PRD  hhhhhhccccceeehhhhhcccceeeeccccccceeeeecccceeeeeeeeeecccccccccc

SEQ  DLSFIFNGDAAPSESFVVLDNEQKVYQRIHHEESEMETEEEVDILMSSDIYSATLSTKSI
SEG  ...............................xxxxxxxxxx...................
PRD  ceeeeeccccceeeeeeeeeccccceeeehhhhhhhhhhhhhhhhhhcccceeeecccccc

SEQ  SFTRAQTGWLFREDKTERVGNFLADFYLVNGLVIESRKRREHLSEEDILRNKAIMESLSK
SEG  ............................................................
PRD  eeeeccccceeeeccccchhhhhhheeeeeeeeeeeeeehhhhhhhhhhhhhhhhhhhhhhhc

SEQ  GGNIMEQNFEPIRRQSLTPPPQNTITWEEYISAENGKAPHLGRELVCKESKKTFKATIAM
SEG  ............................................................
PRD  cceeecccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhh

SEQ  SQEFPLGIELLLNVLEVVAPFKHFNKLREFVQMKLPPGFPVKLDIPVFPTITATVTFQEF
SEG  ............................................................
PRD  hhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccceeeeeeeeeeehhhhhhhcc

SEQ  RYDEFDGSIFTIPDDYKEDPSRFPDL
SEG  ..........................
PRD  ccccccccceeecccccccccccccc
```

Prosite for DKFZphfkd2_46d13.1

PS00001	82->86	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	373->377	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	8->11	PKC_PHOSPHO_SITE	PDOC00005
PS00005	296->299	PKC_PHOSPHO_SITE	PDOC00005
PS00005	316->319	PKC_PHOSPHO_SITE	PDOC00005
PS00005	336->339	PKC_PHOSPHO_SITE	PDOC00005
PS00005	410->413	PKC_PHOSPHO_SITE	PDOC00005
PS00005	413->416	PKC_PHOSPHO_SITE	PDOC00005
PS00006	16->20	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	274->278	CK2_PHOSPHO_SITE	PDOC00006
PS00006	278->282	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	386->390	CK2_PHOSPHO_SITE	PDOC00006
PS00006	476->480	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00008	35->41	MYRISTYL	PDOC00008
PS00008	46->52	MYRISTYL	PDOC00008
PS00008	108->114	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	487->493	MYRISTYL	PDOC00008
PS00016	239->242	RGD	PDOC00016

(No Pfam data available for DKFZphfkd2_46d13.1)

DKFZphfkd2_46j20
----

group: metabolism

DKFZphfkd2_346j20 encodes a novel 224 amino acid protein similar to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.

The new protein seems to be the human ortholog of 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.

The new protein can find application in modulating the homoprotocatechuate degradative pathway and as a enzyme for biotechnologic production processes.

strong similarity to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase

complete cDNA, complete cds, EST hits,
potential start at Bp 16 matches kozak consensus ANCatgG
strong similarity to proteins of worm plant archea and bacteria
2-hydroxyhepta-2,4-diene-1,7-dioate isomerase is part of
the tyrosine metabolism (degradation of tyrosine late step) EC 5.3.1.-
complete cds according to similar C.elegans and A.thaliana protein

Sequenced by MediGenomix

Locus: unknown

Insert length: 1706 bp
Poly A stretch at pos. 1686, polyadenylation signal at pos. 1667

```
   1 CACTTGATGG GAATCATGGC AGCATCCAGG CCATTGTCCC GCTTCTGGGA
  51 GTGGGGAAAG AACATCGTCT GCGTGGGGAG GAACTACGCG GACCACGTCA
 101 GGGAGATGCG CAGCGCGGTG TTGAGCGAGC CCGTGCTGTT CCTGAAGCCG
 151 TCCACGGCCT ACGCGCCCGA GGGCTCGCCC ATCCTCATGC CCGCGTACAC
 201 TCGCAACCTG CACCACGAGC TGGAGCTGGG CGTGGTGATG GGCAAGCGCT
 251 GCCGCGCAGT CCCCGAGGCT GCGGCCATGG ACTACGTGGG CGGCTATGCC
 301 CTGTGCCTGG ATATGACCGC CCGGGACGTG CAGGACGAGT GCAAGAAGAA
 351 GGGGCTGCCC TGGACTCTGG CGAAGAGCTT CACGGCGTCC TGCCCGGTCA
 401 GCGCGTTCGT GCCCAAGGAG AAGATCCCTG ACCCTCACAA GCTGAAGCTC
 451 TGGCTCAAGG TCAACGGCGA ACTCAGACAG GAGGGTGAGA CATCCTCCAT
 501 GATTTTTTCC ATCCCCTACA TCATCAGCTA TGTTTCTAAG ATCATAACCT
 551 TGGAAGAAGG AGATATTATC TTGACTGGGA CGCCAAAGGG AGTTGGACCG
 601 GTTAAAGAAA ACGATGAGAT CGAGGCTGGC ATACACGGGC TGGTCAGTAT
 651 GACATTTAAA GTGGAAAAGC CAGAATATTG AGTTATTTCT TAACAAGTTT
 701 CGAGAGAGAA GGGAGCAAGA CAAGAGCAAG CAACGGCTAT TAAATGTCAC
 751 AATCCTTTAA TTAGAAACCA TTTATTGGCC GGACGCGGTG GCTCACGCCT
 801 GTAATCGCAG CACTTTGGGA GGCCGAGGCG GGCGGCTCAC GACGTCAGGA
 851 GATCCAGACC ATCTTGGCTA ACAGGGTGAA ACCCCGTCTC TACTAAAAAT
 901 ACAAAAAATT AGCCGGGCGT GGTGGCGGGC GCCTGTAGTC CCAGCTACTC
 951 TGGAGGCTGA GGCAGGAGAA TCAATTGAAC CCGGGAGGCG GAGCTTACAG
1001 TGAGCTGAGA TTGCGCCACT GTACTCCTGG GCAACAGCGA GACTCCGTCT
1051 CAAAAAAAAA AAAAAAAAAA AGAAACCATT TATTTTAAAA ATGATTAGAT
1101 TGCTATGCCT CAACTCATAG AAGATGAACC CTTCAAGAAA ACGTGAAGTA
1151 GAACGGGTGG GCCAGAAATG AAAACAGGCA AGTAAAGTAT TTCTTCGGAA
1201 AACATTTTAT CAAACCAAAT GTTAAAAAGA CTTTCCTTTT GTAAAACTGG
1251 ATTAGAGAAG ACTTTTCAGT GGGTTATCTC TAGGATGATC AGTAGTTCAG
1301 CACTTAAAAA CTGCAGAGAA AACTGAAAGT TATGTTCCAG ATAACTTTCC
1351 GTTGTTTACC AAATTTTCTT AGATTTGGTC ATCATCAGGA AGCATTTGTA
1401 AAAATAAAAA TCTCCACAAA TTACTGGCCC ATCTCGGACT TGCTGAATCA
1451 ATTTGATAGG ATTAATCTCC AGTGAAGCTG TGTTTACAGG GCATTCCAAG
1501 TGATTCTTAT CAGGAAATGT GAAAAACACT CCTGTACATA ATCGGTTAAT
1551 TTAAAATTTT ACTTAATAAG TGAACAAGTA ATGAAGATTT CACCTGTTTA
1601 CTTAGGGTAT CTACCCAGAC CCATCGATTC TGAGTTCGGG AGATGATTTT
1651 GAAATTACTG TTTTCCAAAT AAAGGTGCTC CCTTCCAAAA AAAAAAAAAA
1701 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

94039092: Purification, nucleotide sequence and some properties of a bifunctional isomerase/decarboxylase from the homoprotocatechuate degradative pathway of Escherichia coli C.

Peptide information for frame 1

ORF from 7 bp to 678 bp; peptide length: 224
Category: strong similarity to known protein

```
  1 MGIMAASRPL SRFWEWGKNI VCVGRNYADH VREMRSAVLS EPVLFLKPST
 51 AYAPEGSPIL MPAYTRNLHH ELELGVVMGK RCRAVPEAAA MDYVGGYALC
101 LDMTARDVQD ECKKKGLPWT LAKSFTASCP VSAFVPKEKI PDPHKLKLWL
151 KVNGELRQEG ETSSMIFSIP YIISYVSKII TLEEGDIILT GTPKGVGPVK
201 ENDEIEAGIH GLVSMTFKVE KPEY
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_46j20, frame 1

PIR:S44919 ZK688.3 protein - Caenorhabditis elegans, N = 1, Score = 537, P = 8.7e-52

PIR:D71109 probable 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase - Pyrococcus horikoshii, N = 1, Score = 529, P = 6.1e-51

PIR:C71425 hypothetical protein - Arabidopsis thaliana, N = 1, Score = 519, P = 7e-50

PIR:A64864 probable 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase b1180 - Escherichia coli, N = 1, Score = 474, P = 4.1e-45

>PIR:S44919 ZK688.3 protein - Caenorhabditis elegans
Length = 214

HSPs:

Score = 537 (80.6 bits), Expect = 8.7e-52, P = 8.7e-52
Identities = 99/211 (46%), Positives = 138/211 (65%)

```
Query:   10 LSRFWEWGKNIVCVGRNYADHVREMRSAVLSEPVLFLKPSTAYAPEGSPILMPAYTRNLH 69
            L+ F      IVCVGRNY DH  E+ +A+  +P+LF+K   ++  EG PI+ P    +NLH
Sbjct:    4 LAGFRNLATKIVCVGRNYKDHALELGNAIPKKPMLFVKTVNSFIVEGEPIVAPPGCQNLH 63

Query:   70 HELELGVVMGKRCRAVPEAAAMDYVGGYALCLDMTARDVQDECKKKGLPWTLAKSFTASC 129
             E+ELGVV+ K+   + ++ AMDY+GGY + LDMTARD QDE KK G PW LAKSF  SC
Sbjct:   64 QEVELGVVISKKASRISKSDAMDYIGGYTVALDMTARDFQDEAKKAGAPWFLAKSFDGSC 123

Query:  130 PVSAFVPKEKIPDPHKLKLWLKVNGELRQEGETSSMIFSIPYIISYVSKIITLEEGDIIL 189
            P+  F+P   IP+PH ++L+ K+NG+ +Q    T  MIF IP ++ Y ++  TLE GD++L
Sbjct:  124 PIGGFLPVSDIPNPHDVELFCKINGKDQQRCRTDVMIFDIPTLLEYTTQFFTLEVGDVVL 183

Query:  190 TGTPKGVGPVKENDEIEAGIHGLVSMTFKVE 220
            TGTP GV  +    D IE G+    ++  F V+
Sbjct:  184 TGTPAGVTKINSGDVIEFGLTDKLNSKFNVQ 214
```

Pedant information for DKFZphfkd2_46j20, frame 1

Report for DKFZphfkd2_46j20.1

[LENGTH]	224
[MW]	24843.07
[pI]	6.96
[HOMOL]	PIR:S44919 ZK688.3 protein - Caenorhabditis elegans 8e-55
[FUNCAT]	r general function prediction [M. jannaschii, MJ1656] 9e-40
[FUNCAT]	99 unclassified proteins [S. cerevisiae, YNL168c] 4e-38
[EC]	5.3.3.10 5-Carboxymethyl-2-hydroxymuconate delta-isomerase 1e-35
[PIRKW]	isomerase 1e-35
[PIRKW]	intramolecular oxidoreductase 1e-35
[SUPFAM]	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase 1e-46
[PROSITE]	MYRISTYL 4
[PROSITE]	AMIDATION 1

[PROSITE]	CK2_PHOSPHO_SITE	2
[PROSITE]	PKC_PHOSPHO_SITE	3
[KW]	Alpha_Beta	

```
SEQ   MGIMAASRPLSRFWEWGKNIVCVGRNYADHVREMRSAVLSEPVLFLKPSTAYAPEGSPIL
PRD   cccccccccchhhhhhcceeeeeecchhhhhhhhhcccccceeeecccccccccccccc

SEQ   MPAYTRNLHHELELGVVMGKRCRAVPEAAAMDYVGGYALCLDMTARDVQDECKKKGLPWT
PRD   ccccchhhhhhheeecccccccccchhhhhhhheeeeeecccchhhhhhhhhhhcccccc

SEQ   LAKSFTASCPVSAFVPKEKIPDPHKLKLWLKVNGELRQEGETSSMIFSIPYIISYVSKII
PRD   cccccccccccceeeeccccccccccceeeeecccccccccccccceeechhhhhhhhhhh

SEQ   TLEEGDIILTGTPKGVGPVKENDEIEAGIHGLVSMTFKVEKPEY
PRD   hcccccceeeecccccccccccccceeeeeecccccccccccccccc
```

Prosite for DKFZphfkd2_46j20.1

PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	192->195	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00008	2->8	MYRISTYL	PDOC00008
PS00008	75->81	MYRISTYL	PDOC00008
PS00008	116->122	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00009	78->82	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfkd2_46j20.1)

DKFZphfkd2_46k19
----

group: transcription factors

DKFZphfkd2_46k19.3 encodes a novel 130 amino acid protein similar to rat Dcoh, a bifunctional protein-binding transcriptional co-activator.

Dcoh is a bifunctional protein, complexed with biopterin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the biopterin cofactor of phenylalanine hydroxylase.

The new protein can find application in modulating/blocking the expression of genes controlled by the hepatocyte nuclear factor-1.

strong similarity to pterin-4-alpha-carbinolamine dehydratase

potential start at Bp 102 according to similar proteins,
both genomic sequences are from chromosome 5,

Sequenced by MediGenomix

Locus: map="5"

Insert length: 5641 bp
Poly A stretch at pos. 5617, polyadenylation signal at pos. 5598

```
   1 CAGCCCTCGG CAGACGGCCA ATGGCGGCGG TGCTCGGGGC GCTCGGGGCG
  51 ACGCGGCGCT TGTTGGCGGC GCTGCGAGGC CAGAGCCTAG GGCTAGCGGC
 101 CATGTCATCA GGTACTCACA GGTTGATTGC AGAGGAGAGG AACCAAGCTA
 151 TACTTGACCT TAAAGCAGCA GGATGGTCGG AATTAAGTGA GAGAGATGCC
 201 ATCTACAAAG AATTCTCCTT CCACAATTTT AATCAGGCAT TTGGCTTTAT
 251 GTCCCGAGTT GCCCTACAAG CAGAGAAGAT GAATCATCAC CCAGAATGGT
 301 TCAATGTATA CAACAAGGTC CAGATAACTC TCACCTCACA TGACTGTGGT
 351 GAACTGACCA AAAAAGATGT GAAGCTGGCC AAGTTTATTG AAAAAGCAGC
 401 TGCTTCTGTG TGATTTCTTC CAAAATACAT AAGTCTGAGA GGCTAAACTT
 451 GATGGCTGTG TTAACATATG TCACGTGTAG CACAGTGGAG AAAGCAGGAT
 501 ATGGCTCATA ATGACAGTGG TGAAGACCTG CGAATGAAGT TGCTAGTTAA
 551 CACCTACATT AGGGTTTGAC ATAGGTCTAT GTTATGGGTC GCTGCATCTG
 601 CTGGAACTCA CAGACTTTAC TATAGAGAAT CAAAGATCCC GTATCCGAAG
 651 TCTATGGAAA TGCTCATGGT GGTAAATTCC AACAGAATGA AACACCAAAC
 701 TTGCTTAAAG TAACTCACGT TTCAATTTGA AAGAGATATT GTCAAAATTG
 751 GAGGCCCCCA GGTTCCTGTC TGTTCCAAAT CTTTGCATGA TGACAGTGGT
 801 TTCTCTGATG TGGTAAGCTT TGGCTTTCTT CTGTTTTCTT TCTAAAAGAT
 851 CACTGGAGTA GAGAGGAGTT AAACAGACAT GACCTTTGAC CTCTTGCATG
 901 ACCTCCACAG ATAGCAAACC GGGCCGACAC ATGGTTGACG ATGTCCTTTT
 951 CTACAATGAA GTTAATGAAA GTTCTGAAAA TAGTGATTAC TTTCTGACAT
1001 TGATAGGATT TAGGAAACCT CTGGATAAAT AGCTTAAGCA TGGCTGTTTA
1051 TGTTTTTGCT ATAGACAAAA AGCAGCAGCA TGTACATTGT ATTTGGACAC
1101 AAGCCTGCCT CGGTTAATAT ATTGAACTAT TGGACCACTA GGGTTAGTAG
1151 GGAGCGGTCT GTACACTTTC TGATTCAGCA TTCAGAAACA TTCTAGGTGG
1201 ACTCTGTAGC TTTCAGTTTT GTAAAGTTAT CGGAAAAACA TCGGGAGGGT
1251 TTGGCCATCA TATGTGAGCT TTGTGTTTCA ATGCCAGTTA CTCAGGATTA
1301 GTAAATTAAT GACTGTCCAG AGGACTTCAG GGTCACCAAG CTGCTGCACC
1351 TGCCATTGGC TGACTCTCCC CGGCTATCTG TGGCTGAGAT GGTGCTGCTT
1401 AGGTCACGCA GAGCATGAGC TGCTGCTGAA AGGGCACAGG AGATGGCCCT
1451 TGGGCTTCTC ATCCCAGGAT GCCTGCCCTG CCCACCAATC CATGAGAAGA
1501 TATGTATGAT TTCAGTAGGC CCTGGATCAG CTTGTCACCT CTGGTTTCCT
1551 GTTTGCTTTC CACTCACTCA GCTGGAGTTT CATTTCCAGA CTAAAGTCTT
1601 CATCATTGGC TTCAGAAACA GCATTCATCT GTGGCTGTGC TGATGTAGTA
1651 CACCAAGAAC AACTGGGCTC TTCTCTGTCA CTTTCAGTGG GCTACCTTCC
1701 CTCACCTCTC CAAGCAGCAT GAAAGAATTC TTTACATTTT TAATCTCTTT
1751 TTTGTTTTTC CCTGAAAGTA TGCTTTGGTG CTTAAAGAGA GAAGTCACAA
1801 AAGTATACTA CTGAGTTTCC TGGAGATGAA ATCCTGTTGT CCCTAGCTAT
1851 GTGAATGAGC ACAGGGATCC CTGATGCCAT TATTTTGTAT ATTCATACGG
1901 CACACACTTA CTGAGGGCCT TCTGTGTGCC CTAGGGGATT GAGCACAGTG
1951 ACATATCAGG GCAGGTAGAA ACAGATGGAG AGCTGATGCG GGCTGTCTTA
2001 GAGCAGCTGC CCCAGGAGGC CCCTGTGGAT GGATGTTGGG CAGGAGCCCT
2051 GAGACGTTAG GGGCATATAA CTAAAGGACA TAGCAGGAGT TATAGGAGGA
2101 GCTGATCCCT GAGGGAAACA ATGAAGACGG AGAAGATGGG GCTAAAGTTT
2151 GAATTGTGGG GACATTAATC ACGGTGATTC TTAAAACTTT GCTGTTGATG
2201 ATTTTAAATG GAGAAAATGA GTACGTAAGA TGTTATTTCC CAGTTCAGTA
2251 TATAGGTTGC CCACAAAGTA TTTTCCTACC ATGAATGGTC ATATATACTT
2301 GTTGTAGAAT ACCAGGGACA GCAGAGATGG TGGGGTAGTT ACTTCCTTTT
2351 CTTACAGCCC AAGAACTTTG GTGTCCAGGA GATTGACCAA TTTAGCCACT
2401 GAGCATTTAA TACAACACAG GGCTACCCAG ATCCCACTGT CCTGATTTGC
2451 CCTGAAAGCC AAAGGAGTCA GGAGAAGGTG AGTGGGGTGA ATATATTAAT
2501 CCTGAGAGTT GAACAGAGCA AAAATCCCTA TTACTTTTGT ACTTAAAACA
```

```
2551 TCTCTGCCAC ATGTGCTCAC TCTTTATATT CTGTTTAGGT GGTTTATATG
2601 TGCACATCCC ATCCTATGCC TGCAGTTAGC CAACTCAGGG TTTATATTGC
2651 CTCCTTTCTT TTTTTCTTTT TTTTTTTTTT TTTTAAGAGA TGGGGTCTCG
2701 TTCTGTCATG CAGACTGGAG TGCAGTGGTG TGATCACAGC TCATTGTAAC
2751 CTCCAACGCC TGGACTGAAG TGATCCTCCT GCCTTGGCCT CTCTGGTAGC
2801 TGGGACTACA GGTGCATGCC ACCACACCCA CCTAATTTTT TTTATTTTTA
2851 TTTTTTGTAG AGACAGTCTC ACTATCTTGC TCGGGCTGGT CCTGAACTCC
2901 TGGGCTCAAG TTATCTTGCT GCCTCAGCCT CCCATGGGTA ATCTTTATTT
2951 CCTTTTTTTT TTTTTTTTGG AGATGGAGTT TCGCTCTTGT CGCCCAGGCT
3001 GGAGTGCAAT GGCACGATCT TGGCTCACTG CAGTCTCCAC CTCCTGGGTT
3051 CAGGTGATTC TCCATCCTCG GCCTACTGAG TAGCTGAGAT TACAGGCAAC
3101 TGCCACCATG CGCGGCTAAT TTGTGTATTT TTTTTAGTA AGAGATGGGG
3151 TTTCGCCATG TTGGCCGGAC TGGTCTTAGA CTCCTGACCT CAAGCGACCT
3201 GCCTGCCTTG GCCTCCCAAA GTGCTGGGAT TACAGGCATG AGCCGCTATG
3251 CCTCGTCGCT GATTTTTATT TCTTATTTTT TTTTTAGAGA TGGGGGTCTC
3301 ACTATGCTGC TCAGGCTGAT CTCAAACTCC TGGCCTCAAG TGATCCTCCC
3351 ACCTTAGCCT CCCAAGTTGC TGGGATTATA AGTGTGAGCC ACTATCCCTA
3401 CCTCACTATT ACCTTCTTTG CTTCTCTTGT TTTCTTTTGT TCTAAGTCAA
3451 ACCCATCACA ATCTTTTCTT GTCCTTCCAG GTGTTTTCCA GTGCTGTGCC
3501 CTGGATGTGC TCTCTTTCTC TTAGAGCCCA GAGAACTTGC TTTTCCCCCT
3551 TATATATGAC CCTTAACTTT TTCTAACACA TTATTAAGGG CCTGTGTCTA
3601 TCAGCTGGGG GCACTTCTTG AAGGGAGGGC CTTTGTGTGG TCTGTTTCTA
3651 GTGACTTCCA GCTTTAACCC AGAGCCTCAT GATTGCTGGG TGCCCATAGC
3701 CTTTTTGCTG AATGGAGGCA CTCAGTCTCC TTGGGAAGAG AGAATCCATG
3751 ATAGACCCAC TTGGGAGCTC CCCACTTCAG GGGCCTACAC ACTGGTAATG
3801 CAACAGAATG CCCAAGAGTG ACCTCATAAA GCAAGGATTC CCTTCGTGGC
3851 CCCTTCTCTG CTGCCTCTCA GAATCCAGAC GCTAAGGAAA ATCCCTAAGC
3901 AGAGATTTTC TGTTGGATGC TAAAAGCAAG GAATAAAAGT TGAAAATTTG
3951 GAAAATGTCT CAACACCGTC ACCAGCGCCA CTCGAGAGTC ATTTCTAGTT
4001 CACCAGTTGA CACTACATCG GTGGGATTTT GCCCAACATT CAAGAAATTT
4051 AAGTAAATAT TATCTATCTC CATTGCCTGT TAAGAAATGT GCTAGTAGAA
4101 GTGTGAGGGC AGGGTGTCAG TGTTCTCTCA GCCTCTTCCC TCAGATACTC
4151 GTCTGCTTAC CAAAATAAGT TGCATGTCCT TGACAATCTG GTTTCTATGA
4201 TTGGTGAGGC TGGCATGCTA TTACCTTTAT GTGCCCTGTA GACTTGAATG
4251 ACCAGTTTGA CCAGTTTGAC TGTTAGATAA TCAGAAGGCT TTTCTCTTTT
4301 TTTATAATAG ACCCCATCTC AAATCAGATA ATGAAAAATTA CATATCTTGA
4351 TATATTAGAA AAGTATATAC ATTCTGGCTG GGCACGGTGG CTCACGCCTG
4401 TAATCCCTGC ACTTTGAGAG GCTGGGGCGG ATCACTTGAG GTCAGGAGTT
4451 TGAGACCGGC CTGGCCAGCG TGGCGAAACC CCATCTCTAC TAAAAATACA
4501 CAGATTAGCC CGGAGTGATG GTGTGCACCT GTTGTCCCAG CTACTCAGGA
4551 TGCTGAGGCA GGAGAATCCC TTTAACCTGG GGGGCGAAGG TTGCAGTGAG
4601 CCAGGATTGC ACCACTGCAC TCCAGCCTGG GTGACGGAAC GGGACTCTGT
4651 CTCAGAAAAA AAAAAAAAGA AGAGGAAAAA GAAAAATATA TATTCTATAT
4701 TTTTTTAACT TATGAGAATG TGTTCATTTC ATTTGTAACA TATAATGGGA
4751 AACAGTAATA CGTACTCTGA GAAAAATTGC AAAGCACAGA TAAATGGAAA
4801 TAAACAGGAA AAAGAATCAC CTATAACCTC ACCATCCATA GACAGACACT
4851 GTTAAAATTT TGGCATATTT CCTGCTGATT TTTTCTACTG CTGATTTTTG
4901 CACAGGTGAG ATAATTTTGA ACAGAGAATT TTGTATCTTT GGTTTTTGTG
4951 TTTCGCTGCA CACAAAAACA AAAGATATAA AAATGGATCA TAAACATTTT
5001 TCTAAATCCT GAAAAGTGCA TAGACATATT TTAGTGCCTG TATTTCACAA
5051 GATGGACATA CCATAATTTA CTTACACAGT CCTTTTTGTT AGATGTTTAA
5101 GTTGTTTTCA AGCTTCTCAG TGCTGGAAAA AATACTGAGA TAGACATGTT
5151 TAGTTGAAGT TATTTCATTT CAGGTTATAT TATCTTGGGT CAGAGAATGA
5201 ATGGTTCTCA GGCTTTTCAA AAGAGCTGGT CAGTTTTTAT GCCTCTGGCA
5251 GTTTTTGAGA GTGCTCAATC ATACTACACT GTTGCCAGCA TTAGATCTTA
5301 TCACATTTAA GTCATTGCTA ATTTTATAAA CAAAAACAAT GGTTTTACTT
5351 TGCATCTCCC TGATTGGTGT TGCTGTAGAA CATATTTGGA GAAGTTTGTT
5401 TGTCTTTGGT GTTTATTCCA TGAATAGATT GTGTGCCCAT TTTCTCTTGG
5451 GGTATTCAGT TTTTTATTAC TGATGTGAGC ATGTGTATGG GTGATTATTT
5501 GATGATTATC AGTTTTGCTT AGTAGACTGG CAATATTTAG TCTTGCTGTC
5551 ACTGTGTTCC CAGTGCCAAC TAGATTGCTT GATATGTAGT TGCCACTCAA
5601 TAAAGATTTG TTGAGTCAAT GAAAAAAAAA AAAAAAAAAA A
```

BLAST Results

Entry AC004764 from database EMBL:
Homo sapiens chromosome 5, P1 clone 255g5 (LBNL H61), complete sequence.
Score = 11057, P = 0.0e+00, identities = 2217/2224
Bp 428-5625 of cDNA == Bp 2912-8107 of AC004764

Entry HSAC1555 from database EMBL:
Homo sapiens (subclone 1_d8 from BAC H75) DNA sequence, complete sequence.
Score = 575, P = 5.1e-30, identities = 115/115
Bp ~240- 430 of cDNA == HSAC1555 splice pattern

Medline entries

93186787:
Phenylalanine hydroxylase-stimulating protein/pterin-4 alpha-carbinolamine dehydratase from rat and human liver. Purification, characterization, and complete amino acid sequence.

93101632:
Identity of 4a-carbinolamine dehydratase, a component of the phenylalanine hydroxylation system, and DCoH, a transregulator of homeodomain proteins.

95242099:
Crystal structure of DCoH, a bifunctional, protein-binding transcriptional coactivator

Peptide information for frame 3

ORF from 21 bp to 410 bp; peptide length: 130
Category: strong similarity to known protein

```
  1 MAAVLGALGA TRRLLAALRG QSLGLAAMSS GTHRLIAEER NQAILDLKAA
 51 GWSELSERDA IYKEFSFHNF NQAFGFMSRV ALQAEKMNHH PEWFNVYNKV
101 QITLTSHDCG ELTKKDVKLA KFIEKAAASV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_46k19, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_46k19, frame 3

Report for DKFZphfkd2_46k19.3

```
[LENGTH]     130
[MW]         14377.56
[pI]         9.17
[HOMOL]      PIR:A47189 pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96) - rat 4e-34

[FUNCAT]     01.07.99 other vitamin, cofactor, and prosthetic group activities  [S.
cerevisiae, YHL018w] 5e-04
[SCOP]       d1dchg_ 4.38.1.1.1 Pterin-4a-carbinolamine dehydratas 4e-50
[EC]         4.2.1.96 Tetrahydrobiopterin dehydratase 6e-34
[PIRKW]      nucleus 6e-34
[PIRKW]      carbon-oxygen lyase 6e-34
[PIRKW]      homotetramer 6e-34
[PIRKW]      hydro-lyase 6e-34
[PIRKW]      cytosol 6e-34
[PIRKW]      acetylated amino end 6e-34
[PIRKW]      homodimer 6e-34
[SUPFAM]     pterin-4-alpha-carbinolamine dehydratase 6e-34
[PROSITE]    MYRISTYL         2
[PROSITE]    CK2_PHOSPHO_SITE     3
[PROSITE]    PKC_PHOSPHO_SITE     4
[KW]         Alpha_Beta
[KW]         3D
[KW]         LOW_COMPLEXITY    14.62 %
```

```
SEQ    MAAVLGALGATRRLLAALRGQSLGLAAMSSGTHRLIAEERNQAILDLKAAGWSELSERDA
SEG    .xxxxxxxxxxxxxxxxxxx........................................
1dchB  ..............................CCCCHHHHHHHHHHHHHHCCEEECCCCE

SEQ    IYKEFSFHNFNQAFGFMSRVALQAEKMNHHPEWFNVYNKVQITLTSHDCGELTKKDVKLA
SEG    ............................................................
1dchB  EEEEEECCCHHHHHHHHHHHHHHHHHHCCCCEEEETTTEEEEEECBTTTTBTCCHHHHHH

SEQ    KFIEKAAASV
```

```
SEG      ..........
1dchB    HHHHHHHHHH
```

Prosite for DKFZphfkd2_46k19.3

PS00005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS00005	32->35	PKC_PHOSPHO_SITE	PDOC00005
PS00005	56->59	PKC_PHOSPHO_SITE	PDOC00005
PS00005	113->116	PKC_PHOSPHO_SITE	PDOC00005
PS00006	56->60	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	113->117	CK2_PHOSPHO_SITE	PDOC00006
PS00008	6->12	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfkd2_46k19.3)

DKFZphfkd2_46m4
----

group: signal transduction

DKFZphfkd2_46m4.3 encodes a novel 198 amino acid putative GTP-binding protein related to the SAR-1 family of Ras superfamily members.

SAR1 proteins are involved in vesicular transport between the endoplasmic reticulum and the Golgi apparatus.

The new protein can find clinical application in modulating the transport of vesicles to the Golgi Apparatus, thus enabling post-translational modifications of the vesicles contents. Blocking of the molecule is expected to result modulation/blocking of secretory pathways.

nearly identical to mouse GTP-binding protein

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="438.9 cR from top of Chr10 linkage group"

Insert length: 2996 bp
Poly A stretch at pos. 2969, polyadenylation signal at pos. 2958

```
   1 ACATCCGGCG AGTAGCTGGC GGTCCCGGGT GCTGCTGGTT AGTGTGCTCT
  51 GAGGGAGGGT CCGAGCCAGC CGCTGTTTTG CCGGAGGAGC CCCTCAGGCC
 101 GTAGTAAGCA TTAATAATGT CTTTCATCTT TGAGTGGATC TACAATGGCT
 151 TCAGCAGTGT GCTCCAGTTC CTAGGACTGT ACAAGAAATC TGGAAAACTT
 201 GTATTCTTAG GTTTGGATAA TGCAGGCAAA ACCACTCTTC TTCACATGCT
 251 CAAAGATGAC AGATTGGGCC AACATGTTCC AACACTACAT CCGACATCAG
 301 AAGAGCTAAC AATTGCTGGA ATGACCTTTA CAACTTTTGA TCTTGGTGGG
 351 CACGAGCAAG CACGTCGCGT TTGGAAAAAT TATCTCCCAG CAATTAATGG
 401 GATTGTCTTT CTGGTGGACT GTGCAGATCA TTCTCGCCTC GTGGAATCCA
 451 AAGTTGAGCT TAATGCTTTA ATGACTGATG AAACAATATC CAATGTGCCA
 501 ATCCTTATCT TGGGTAACAA AATTGACAGA ACAGATGCAA TCAGTGAAGA
 551 AAAACTCCGT GAGATATTTG GGCTTTATGG ACAGACCACA GGAAAGGGGA
 601 ATGTGACCCT GAAGGAGCTG AATGCTCGCC CCATGGAAGT GTTCATGTGC
 651 AGTGTGCTCA AGAGGCAAGG TTACGGCGAG GGTTTCCGCT GGCTCTCCCA
 701 GTATATTGAC TGATGTTTGG ACGGTGAAAA TAAAAGAGTT TTACTTCTCT
 751 GGACTGATCC TATTCACAGC TTCCTCATGA ACTTTTCTAA TAGAACAAGG
 801 ATAGCTCTCC AACCATGTCT GGCGTTGAGA AGCCAAGAGT CTCTGTCAAC
 851 TCTCTCATTG CCCAGTGGTG ACATGTGCTC TTCTCCACAC TGTTGGGAGG
 901 TAATGCTGCC CCACGTGCTG GTGCAGGTCA GTATCCTGGG ACTTGGAAGC
 951 TGGCAGGATT TGCCGGGTAA AGCTGTATGC CATCATGGGG CACCTGAAAA
1001 GAAAAACACG TCTCACCACT GTGGTTGATT CAAAAGAAAG TGATTCTATT
1051 TTTTAAAGAA AGCGTTGTTA ATGTAATTGG TATCCCTCCT AACTTTTTGA
1101 GTTCACAATT TACTTGGTCC AGAGTTTTCT ATTCTTTTTT TTTTTTTAAA
1151 CTAATGAATG ACATTTAGAT ACTTCATAAA ATTATGAACA GATATGGAGG
1201 CCAGAGCTCA TTTGGGTAAA CTTACTCCTG CTGAGTTAGC AGGTTGGTGA
1251 GAGAAGCTCC CCTGAGCTCA CCTGTCTCTC TGACTGCCTT GGAGTAGGTG
1301 GCATAACCTT GTGCACAGAG AACTAGAAAA GGGGCAGAAC CCCGGCCTTG
1351 CAGTTGTGGC AGGTTTCCAC TGTGGTAAGC TAGGTTCATT CCTCATCAAG
1401 GAATGTGTAG CAGATTGTTC ACTGTGGAGG AGGTAATTAT AGAATGGGTT
1451 ATTGTTGTTA TTCTTACTCA TGAAGTTACA GATTTTAGCC AGTCTTTGCT
1501 TTTATACTTT TGTGAAATTT AATTTCTCTC TATAGCACCT TCCTTTTTCG
1551 TTTTCAGTTA TCAAAAGTGA CTTTGACCTC ATAAGAGAGT TGAGAACATC
1601 TCTCGTGTCA CATACTGCAG GTGCATCAGT TACTTTTGCA CAGATTCTAG
1651 GGGGACATTT TTCTGAATAG GAAGACAGGA CAAAGTTAAC AGCTTAAGGG
1701 CTCTTAATTC TGTGAGTTGA GGACTTAAAA GTATTGTAGC ATTTGTTTGG
1751 ATCCATGAAA AATGTATTCA GTGGGCTTTA AAATTTCCAT TTGCAGAATT
1801 TGGTCTCTCA GGCTGTTTGG GAGCTCTTTT TTTTACATTT TTTCTCCTTT
1851 GACACCTATT TTATTGGTGT TTAAAGTAAA GGTTAACATC TGTAGCTTTT
1901 CCAGGTTTTT TTTTTTTTTT TTGATATGAA ATTGTCTTTC TCCATTGCAG
1951 AAATAAGCTA GGGAAACACT AACCCAAAAA CTTTCTGTAG AGCTGTTCCT
2001 TTGGAGGCAG CATCACTTAT TGGCAGTAAA GACTCAGTAT AAAAGCACCA
2051 GCATCCCTAC TTGGGTGATG GGGATTAATT TTATAGCATT CCATTTTCCT
2101 AGTGCCACAT GTGAAATTGG ATTTTGATGA TCTTAATCTA TATTCTACCC
2151 TTATAATAAA AGATCAAAAG ATATATCTCC TATGAACAGA TTGCAGATAG
2201 GAGATGAAAA GTTGGGAGGA TGCCTTTATT CTAATGTGAG GGTAGGGAAA
2251 ATGTGGATAA CATTACTGGG GTGAAGGAGG CATTGTTCTT TAGTTGGAGT
2301 TCTCATTTTT ATTCTCCAGT ACTGACTTGT GGGGAAAGCA TACTTTTTCA
2351 CTGCCAGGTA CTGAATGCAG AGGCTCAGTG AAGTATATAT GTGGGAAGTG
2401 CATGCATTTC GTTTATTAGC AAACATAGCT GGATTAAGAC GAAGTTGTTG
2451 GTTTGGAAAG GGGTTAAAGC CTTAAGTGAA CAAATCTAGC TAACAGTGAA
2501 TGAACTAGGT AATATAACTT GCATATTTTT AATTTCCTTT GGTTAAAGGT
2551 CCCCCATACT TCTCTGTTCG GAGACATGAG AAGTATGATT ACTTCAGTGT
```

```
2601 TAGTTTTCTT AATTTTTTTT TTCCCCTATT TGTCCCTTGT CACTTTGTTG
2651 CAAGCTAGAA ATCTGTGGGT TATACATAGG GCAGCTCTTT GCGAAAGTGG
2701 TTTATTCCAC TGGAGAAAGG GGATTGAAAA TCAGTTAGAA CCAATGTATT
2751 TCTTGCCCCA CGGAACACTA TTCCTATAAG ATAGCTGAAA GAAGCTGCTG
2801 TGAGGAGCTC AGCTCCAACA CAGGATCAGC ACCTTGTATA GGAATTCCCA
2851 TGAATTATGA CTTCTCATTC TGTTTTATCA GAGTGCATAT ATGTCCTACT
2901 TCAGGAAAAG TAAAACAGTC ATTTACGAAA GAAAGTCAAT CTGTATCCTA
2951 AGCATTTTAA TAAAAAGTTA AAACAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HS679348 from database EMBL:
human STS WI-16722.
Length = 265
Minus Strand HSPs:
Score = 1242 (186.4 bits), Expect = 2.8e-50, P = 2.8e-50
Identities = 260/265 (98%)

Medline entries

94085558:
Molecular analysis of SAR1-related cDNAs from a mouse pituitary cell line.

Peptide information for frame 3

ORF from 117 bp to 710 bp; peptide length: 198
Category: strong similarity to known protein

```
  1 MSFIFEWIYN GFSSVLQFLG LYKKSGKLVF LGLDNAGKTT LLHMLKDDRL
 51 GQHVPTLHPT SEELTIAGMT FTTFDLGGHE QARRVWKNYL PAINGIVFLV
101 DCADHSRLVE SKVELNALMT DETISNVPIL ILGNKIDRTD AISEEKLREI
151 FGLYGQTTGK GNVTLKELNA RPMEVFMCSV LKRQGYGEGF RWLSQYID
```

BLASTP hits

Entry S39543 from database PIR:
GTP-binding protein - mouse
Length = 198
Score = 1029 (362.2 bits), Expect = 5.1e-104, P = 5.1e-104
Identities = 197/198 (99%), Positives = 198/198 (100%)

Entry SARA_MOUSE from database SWISSPROT:
GTP-BINDING PROTEIN SARA.
Length = 198
Score = 1012 (356.2 bits), Expect = 3.2e-102, P = 3.2e-102
Identities = 195/198 (98%), Positives = 196/198 (98%)

Entry CEZK180_4 from database TREMBL:
gene: "ZK180.4"; Caenorhabditis elegans cosmid ZK180.
Length = 193
Score = 679 (239.0 bits), Expect = 6.3e-67, P = 6.3e-67
Identities = 125/197 (63%), Positives = 161/197 (81%)

Alert BLASTP hits for DKFZphfkd2_46m4, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_46m4, frame 3

Report for DKFZphfkd2_46m4.3

[LENGTH]	198
[MW]	22367.00
[pI]	6.21
[HOMOL]	PIR:S39543 GTP-binding protein - mouse 1e-112

[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YPL218w] 1e-58
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YPL218w] 1e-58
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR094w] 2e-23
[FUNCAT] 06.07 protein modification (glycolsylation, acylation, myristylation, palmitylation, farnesylation and processing) [S. cerevisiae, YPL051w] 4e-22
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YDL192w] 3e-20
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBR164c] 3e-19
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YMR138w] 2e-09
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YMR138w] 2e-09
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YHR168w] 7e-05
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YHR005c] 1e-04
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YKL154w] 1e-04
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YHR005c] 1e-04
[FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YHR005c] 1e-04
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YKL154w] 1e-04
[FUNCAT] 08.19 cellular import [S. cerevisiae, YML001w] 3e-04
[BLOCKS] BL00395A Alanine racemase pyridoxal-phosphate attachment site proteins
[BLOCKS] BL01019B ADP-ribosylation factors family proteins
[BLOCKS] BL01019A ADP-ribosylation factors family proteins
[BLOCKS] BL01020D SAR1 family proteins
[BLOCKS] BL01020C SAR1 family proteins
[BLOCKS] BL01020B SAR1 family proteins
[BLOCKS] BL01020A SAR1 family proteins
[SCOP] d1p1j__ 3.25.1.3.1 cH-p21 Ras protein [human (Homo sapiens) 7e-36
[SCOP] d1guaa_ 3.25.1.3.10 Rap1A [Human (Homo sapiens) 8e-40
[SCOP] d1rrf__ 3.25.1.3.5 ADP-ribosylation factor 1 (ARF1) [rat (Rattu 2e-55
[SCOP] d1hurb_ 3.25.1.3.4 ADP-ribosylation factor 1 (ARF1) [human (Hom 1e-58
[SCOP] d1gota2 3.25.1.3.3 (1-54,171-326) Transducin (alpha subunit) [ra 2e-33
[SCOP] d1tadb2 3.25.1.3.2 (1-30,152-316) Transducin (alpha subunit 6e-36
[PIRKW] glycoprotein 4e-19
[PIRKW] monomer 1e-16
[PIRKW] P-loop 3e-64
[PIRKW] lipoprotein 4e-19
[PIRKW] GTP binding 3e-64
[SUPFAM] ADP-ribosylation factor 5e-22
[PROSITE] ATP_GTP_A 1
[PROSITE] MYRISTYL 3
[PROSITE] SAR1 1
[PROSITE] CK2_PHOSPHO_SITE 4
[PROSITE] PKC_PHOSPHO_SITE 3
[PROSITE] ASN_GLYCOSYLATION 1
[PFAM] ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
[KW] Alpha_Beta
[KW] 3D

```
SEQ    MSFIFEWIYNGFSSVLQFLGLYKKSGKLVFLGLDNAGKTTLLHMLKDDRLGQHVPTLHPT
1hurA  .................TTTTTCCCCEEEEEEETTTTCHHHHHHHHCCCCEEEEEEEETTEE

SEQ    SEELTIAGMTFTTFDLGGHEQARRVWKNYLPAINGIVFLVDCADHSRLVESKVELNALMT
1hurA  EEEEEETTEEEEEEETTTTTTTCCCHHHHHHCEEEEEEEEETTTTTHHHHHHHHHHHHHH

SEQ    DETISNVPILILGNKIDRTDAISEEKLREIFGLYGQTTGKGNVTLKELNARPMEVFMCSV
1hurA  TTTTTTTEEEEEEETTTTTTTCCHHHHHHHHCGG.........................

SEQ    LKRQGYGEGFRWLSQYID
1hurA  ..................
```

Prosite for DKFZphfkd2_46m4.3

PS00001	162->166	ASN_GLYCOSYLATION	PDOC00001
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00006	60->64	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	111->115	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	68->74	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00017	32->40	ATP_GTP_A	PDOC00017
PS01020	171->197	SAR1	PDOC00782

Pfam for DKFZphfkd2_46m4.3

```
HMM_NAME     ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)

HMM               *GMgWfsIFrkMWGlWNKEMRILMLGLDNAGKTTILYMLKlgEIVTTIPT
                    ++ FS++++++GL++K++++++LGLDNAGKTT+L+MLK++++ +++PT
Query          9   -YNGFSSVLQFLGLYKKSGKLVFLGLDNAGKTTLLHMLKDDRLGQHVPT     56

HMM                IGFNVETVeYKNIKFNVWDVGGQdsIRPYWRHYYpNTDGIIWVVDSaDRD
                   +++++E++++ +++F+++D+GG++++R++W++Y P+++GI+++VD+AD++
Query         57   LHPTSEELTIAGMTFTTFDLGGHEQARRVWKNYLPAINGIVFLVDCADHS    106

HMM                RMeEaKqELHaMLNEEELrDAPlLIFANKQDLPgAMSesEIREaLGLHeI
                   R+ E+K+EL+A++++E ++++P+LI++NK+D+ +A+SE+++RE+ GL+ +
Query        107   RLVESKVELNALMTDETISNVPILILGNKIDRTDAISEEKLREIFGLYGQ    156

HMM                RCn...........RPWYIQMCCAVtGEGLYEGMDWLSNYInkRkK*
                   +++           RP++++MC++++++G++EG++WLS+YI
Query        157   TTGKGNVTLKELNARPMEVFMCSVLKRQGYGEGFRWLSQYI-----     197
```

DKFZphfkd2_47a4
----

group: transcription factor

DKFZphfkd2_47a4.1 encodes a novel 280 amino acid protein with similarity to zinc finger proteins.

The new protein is a putative transcription factor with one C2H2 zinc fingers.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

similarity to C.elegans F46B6.7

potential frame shift at 1092, will be checked see BLASTX

Sequenced by MediGenomix

Locus: map="7q31"

Insert length: 1756 bp
Poly A stretch at pos. 1737, no polyadenylation signal found

```
   1 CCCTTTTCTT TTCTGCCGGG TAATGGCTGC TTCCAAGACC CAGGGGGCTG
  51 TCGCCCGAAT GCAGGAAGAC CGTGATGGGA GCTGCAGCAC AGTCGGGGGT
 101 GTAGGTTATG GGGTAAGGAT TGTATCCTGG AGCCGCTTTC CCTGCCAGAA
 151 AGTCCAGGTG GCACCACCAC TTTAGAAGGT TCTCCATCTG TGCCTTGTAT
 201 TTTCTGTGAA GAACATTTTC CTGTGGCTGA ACAAGACAAA CTTCTGAAGC
 251 ACATGATTAT TGAGCATAAG ATTGTCATAG CTGATGTCAA GTTGGTTGCT
 301 GATTTCCAAA GGTACATTTT ATATTGGAGG AAAAGGTTCA CTGAACAGCC
 351 CATCACAGAT TTTTGTAGTG TAATAAGAAT TAATTCCACT GCTCCATTTG
 401 AAGAACAAGA GAATTATTTT TTGTTATGTG ACGTTTTACC AGAAGATAGA
 451 ATTCTTAGAG AAGAGCTTCA GAAACAGAGA CTGAGAGAAA TTCTGGAACA
 501 ACAGCAGCAA GAACGAAATG ATAACAATTT TCATGGCGTT TGTATGTTTT
 551 GCAATGAAGA ATTCCTTGGA AACAGATCTG TTATTTTGAA CCACATGGCC
 601 AGAGAACATG CTTTCAACAT TGGATTGCCA GACAACATTG TAAACTGCAA
 651 TGAATTTTTG TGTACATTAC AGAAAAAGCT TGACAATTTG CAGTGCTTGT
 701 ACTGTGAGAA GACCTTCAGG GGCAAAAATA CACTTAAAGA TCACATGAGG
 751 AAAAAACAGC ATCGTAAGAT TAATCCTAAG AACAGAGAAT ATGACAGATT
 801 TTATGTCATC AATTATTTGG AACTTGGAAA ATCGTGGGAG GAAGTTCAGT
 851 TGGAAGATGA TCGGGAGTTG CTGGACCATC AGGAAGATGA CTGGTCTGAT
 901 TGGGAAGAAC ACCCTGCCTC TGCAGTCTGC TTATTTTGTG AAAAGCAAGC
 951 AGAAACAATT GAGAAGTTGT ATGTCCACAT GGAGGATGCA CACGAATTTG
1001 ATCTTCTCAA AATAAAGTCA GAACTTGGAT TAAATTTCTA TCAGCAAGTG
1051 AAACTGGTCA ATTTTATTCG GAGGCAAGTT CACCAATGCA GATGATGGCT
1101 GCCATGTGAA GTTCAAATCC AAAGCAGACT TAAGAACTCA CATGGAAGAA
1151 ACTAAACACA CTTCGCTGCT CCCCGATAGA AAGACGTGGG ATCAACTGGA
1201 GTATTATTTT CCAACCTATG AAAATGACAC TCTCCTGTGT ACACTATCTG
1251 ACAGTGAAAG TGACCTGACA GCTCAGGAAC AAAATGAAAA TGTTCCCATC
1301 ATCAGTGAAG ATACATCTAA ACTGTATGCT TTGAAACAAA GCAGTATTTT
1351 GAACCAGTTG CTACTATAAG AGTACTTGAA AACCTAGAAG AAACTACCAC
1401 AGAAGCAATT TTTCATGTTT TTCTCCTATG AGACAGATAT GAAAGAACAA
1451 TTTAAATTTG AACATCAACA AAAGATTGGT CCTTGGTGAA ATAAACTTTT
1501 CAAAAATGAA TGTTCTTTTC AAAAAATAAA GTAGAAAAAT GCACTTACTA
1551 AGAACATGAA AAAAAAATGA AGTAGGAAAA TAAGATGAAG ACTTTGTATT
1601 TTGGCTGTAA AGTTTTATTG TGTGATCATC TTAAATTATC TCACTTCATT
1651 AAACTCATAA TTATATATAG AAGTATATGT CAATTACAAA GAAATGAAAT
1701 GTTCAAATTA TTTATAAACC TGATTTTTCA ATCAGCGAAA AAAAAAAAAA
1751 AAAAAA
```

BLAST Results

Entry AC004112 from database EMBL:
Homo sapiens BAC clone RG313E03 from 7q31, complete sequence.
Score = 2660, P = 3.0e-241, identities = 534/535
> 10 exons

Entry AC004111 from database EMBL:
Homo sapiens BAC clone RG103H13 from 7q31, complete sequence.
Score = 598, P = 5.8e-17, identities = 128/137
1 exon

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 253 bp to 1092 bp; peptide length: 280
Category: similarity to unknown protein

```
  1 MIIEHKIVIA DVKLVADFQR YILYWRKRFT EQPITDFCSV IRINSTAPFE
 51 EQENYFLLCD VLPEDRILRE ELQKQRLREI LEQQQQERND NNFHGVCMFC
101 NEEFLGNRSV ILNHMAREHA FNIGLPDNIV NCNEFLCTLQ KKLDNLQCLY
151 CEKTFRGKNT LKDHMRKKQH RKINPKNREY DRFYVINYLE LGKSWEEVQL
201 EDDRELLDHQ EDDWSDWEEH PASAVCLFCE KQAETIEKLY VHMEDAHEFD
251 LLKIKSELGL NFYQQVKLVN FIRRQVHQCR
```

BLASTP hits

Entry CEF46B6_6 from database TREMBLNEW:
product: "F46B6.7"; Caenorhabditis elegans cosmid F46B6
>TREMBL:CEF46B6_6 product: "F46B6.7"; Caenorhabditis elegans cosmid F46B6
Score = 630, P = 1.1e-61, identities = 123/289, positives = 183/289

Entry AF059531_1 from database TREMBLNEW:
gene: "PRMT3"; product: "protein arginine N-methyltransferase 3"; Homo sapiens protein arginine N-methyltransferase 3 (PRMT3) mRNA, partial cds. >TREMBL:AF059531_1 gene: "PRMT3"; product: "protein arginine N-methyltransferase 3"; Homo sapiens protein arginine N-methyltransferase 3 (PRMT3) mRNA, partial cds.
Score = 120, P = 1.5e-04, identities = 23/78, positives = 42/78

Entry YB9M_YEAST from database SWISSPROT:
34.7 KD PROTEIN IN SHM1-MRPL37 INTERGENIC REGION.
Score = 112, P = 4.6e-04, identities = 43/165, positives = 71/165

Alert BLASTP hits for DKFZphfkd2_47a4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_47a4, frame 1

Report for DKFZphfkd2_47a4.1

```
[LENGTH]    280
[MW]        33921.94
[pI]        5.63
[HOMOL]     TREMBL:CEF46B6_5 gene: "F46B6.7"; Caenorhabditis elegans cosmid F46B6 1e-56
[BLOCKS]    BL01032B Protein phosphatase 2C proteins
[BLOCKS]    BL00028 Zinc finger, C2H2 type, domain proteins
[PROSITE]   MYRISTYL           1
[PROSITE]   ZINC_FINGER_C2H2        1
[PROSITE]   CAMP_PHOSPHO_SITE       1
[PROSITE]   CK2_PHOSPHO_SITE        3
[PROSITE]   TYR_PHOSPHO_SITE        2
[PROSITE]   PKC_PHOSPHO_SITE        2
[PROSITE]   ASN_GLYCOSYLATION       2
[PFAM]      Zinc finger, C2H2 type
[KW]        Alpha_Beta
[KW]        LOW_COMPLEXITY     8.21 %
```

```
SEQ  MIIEHKIVIADVKLVADFQRYILYWRKRFTEQPITDFCSVIRINSTAPFEEQENYFLLCD
SEG  ............................................................
PRD  cccccceeehhhhhhhhhhhhhhhhhhhhhhhccccceeeeeeccccccchhhhheeeecc

SEQ  VLPEDRILREELQKQRLREILEQQQQERNDNNFHGVCMFCNEEFLGNRSVILNHMAREHA
SEG  .....xxxxxxxxxxxxxxxxxxxxxxx................................
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhccccceeeeeecccccccccceeeehhhhhhhh

SEQ  FNIGLPDNIVNCNEFLCTLQKKLDNLQCLYCEKTFRGKNTLKDHMRKKQHRKINPKNREY
```

```
SEG     ............................................................
PRD     hccccccccccchhhhhhhhhhhhhhhhheeecccccccchhhhhhhhhhhccccccccc

SEQ     DRFYVINYLELGKSWEEVQLEDDRELLDHQEDDWSDWEEHPASAVCLFCEKQAETIEKLY
SEG     ............................................................
PRD     ceeeeeeeecccchhhhhhhhcchhhhhhccccccccccccccchhhhhhhhhhhhh

SEQ     VHMEDAHEFDLLKIKSELGLNFYQQVKLVNFIRRQVHQCR
SEG     ........................................
PRD     hhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhcccc
```

Prosite for DKFZphfkd2_47a4.1

PS00001	44->48	ASN_GLYCOSYLATION	PDOC00001
PS00001	107->111	ASN_GLYCOSYLATION	PDOC00001
PS00004	27->31	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00005	160->163	PKC_PHOSPHO_SITE	PDOC00005
PS00006	160->164	CK2_PHOSPHO_SITE	PDOC00006
PS00006	194->198	CK2_PHOSPHO_SITE	PDOC00006
PS00006	215->219	CK2_PHOSPHO_SITE	PDOC00006
PS00007	178->185	TYR_PHOSPHO_SITE	PDOC00007
PS00007	13->22	TYR_PHOSPHO_SITE	PDOC00007
PS00008	124->130	MYRISTYL	PDOC00008
PS00028	148->171	ZINC_FINGER_C2H2	PDOC00028

Pfam for DKFZphfkd2_47a4.1

```
HMM_NAME      Zinc finger, C2H2 type

HMM               *CpwPDCgKtFrrwsNLrRHMR..T.H*
                   C +  C+KTFR + +L+ HMR    H
Query         148  CLY--CEKTFRGKNTLKDHMRKK-QH     170
```

DKFZphfkd2_4b6
----

group: kidney derived

DKFZphfkd2_4b6 encodes a novel 133 amino acid protein with similarity to Homo sapiens clone 25003 partial CDS.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to Homo sapiens clone 25003

complete cDNA, complete cds, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1936 bp
Poly A stretch at pos. 1916, polyadenylation signal at pos. 1890

```
   1 GGGAGACTTG CAATGAAGTT AGAATGAACA GGAGGAGTCT GCAGCTTTTC
  51 AGTGCCTGGG ATAACTATAG TTTAAAGATC ATTGTGTAAA ATAGGATTTT
 101 TAGTCAGCAT GCATTGTTTT AAACCGACTA ACTGATAGCC TAAAACTTTA
 151 TTTTTGCATT TTGCCAATCC TTGGAGTTTT GTTTTGCAGA ATTAAGAAAA
 201 AAATGAATGT ATGATCATCT GAAAAGGGCT TTCTCTCAAT CCCACTTCAT
 251 GGCATGACCT CTGCTGGATC ATTAGTTCTA GCCAGAGAAG TAGCAAAGGA
 301 ACATGACGTC TGAGACCTCC CTTCCCTCAT CAGTGGGGCT GACTGAGCTG
 351 GGGGCTTGAA GCCGGAGGTA ACCTTTCCTG TCGAATGTTT CTTTAGAGAA
 401 TGGCAATGGT CTCTGCGATG TCCTGGGTCC TGTATTTGTG GATAAGTGCT
 451 TGTGCAATGC TACTCTGCCA TGGATCCCTT CAGCACACTT TCCAGCAGCA
 501 TCACCTGCAC AGACCAGAAG GAGGGACGTG TGAAGTGATA GCAGCACACC
 551 GATGTTGCAA CAAGAATCGC ATTGAGGAGC GGTCACAAAC AGTAAAGTGT
 601 TCCTGTCTAC CTGGAAAAGT GGCTGGAACA ACAAGAAACC GGCCTTCTTG
 651 CGTCGATGCC TCCATAGTGA TTTGGAAATG GTGGTGTGAG ATGGAGCCTT
 701 GCCTAGAAGG AGAAGAATGT AAGACACTCC CTGACAATTC TGGATGGATG
 751 TGCGCAACAG GCAACAAAAT TAAGACCACG AGAATTCACC CAAGAACCTA
 801 ACAGAAGCAT TTGTGGTAGT AAAGGAAAAC CAACCCTCTG GAAAATACAT
 851 TTTGAGAATC TCAAACATCT CACATATATA CAAGCCAAAT GGATTTCTTA
 901 CTTGCACTTT GACTGGCTAC CAGATAATCA CAGTGCGTTT AGTGTGTGTA
 951 ACGAAATATC CTACAGTGAG AAGACACAGC GTTTTGGCAT CACCATGGAA
1001 AGTGGGCTTA AAAAAGGGTC TTCTCAGTGA AATTTTTGGG CATCATGAAG
1051 AACGATCAAC TATCTTCTAA TTTGAATCTA TAGTTACTTT GTACCATTTG
1101 AAATATATGT ATATATATAT ATATAATATT TTGAAATATT ATCTATTCTC
1151 TTCAAGAAAT GAACAGTACC ACAGTTTGAG ACGGCTGGTG TACCCCTTTG
1201 AGTTTTGGAT GTTTTGTCTG TTTTGCTTTG TTTTGTTAGT CATTTCTTTT
1251 TCTAACGGCA AGGAAGATAT GTGCCCTTTT GAGAATTCAA GATGGCACTG
1301 ACACGGGAAG GCCAGCTACA GGTGGACTCC TGGAATTTGA GGCATCATAA
1351 TGATACTGAA TCAAGAACTT CCTTCTGCTT CTACCAGATG GCCCAAGGAA
1401 GCACATCGTC CTGTTTTATT GCTTTCTACC CTGTGCAATA TTAGCATGCA
1451 AGCTTGGCTT ACATAGTCAT ACTTTATATT CAATTGATAT ATAATAACCG
1501 TTCTAACCTC TTCCAGGAAA ATATTTTTAG AACTACTAGC TTTTCCACTT
1551 AGAAGAAAAT GAGGATTCTT AAGGGAGCCA CTCCACCATG CTATTAAGAC
1601 TCTGGCAGAG TTATGGGTAG GATATGGATC CCTACATGAA TAAGTCCTGT
1651 AAATACAATG TCTTAAGGCT TTGTATAGCT GTCCTAGACT GCAGAAATGT
1701 CCTCTGATTA AATCCAAAGT CTGGCATCGT TAACTACATA GTGCTGTAGC
1751 AACAAGTCTT ATCATGGCAT CTCTTTCTAT GTTTGGTTTG CTTTTTCCAA
1801 GAGTATTCAG GTCTCCTCTT GTGAGATAGG AAGGCCATGA AAACAATTAG
1851 ATTTCAAGAT GATCTATGTG ACCAAATGTT GGACAGCCCT ATTAAAGTGG
1901 TAAACAACTT CTTTCTAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

```
ORF from 400 bp to 798 bp; peptide length: 133
Category: similarity to unknown protein
Classification: no clue

    1 MAMVSAMSWV LYLWISACAM LLCHGSLQHT FQQHHLHRPE GGTCEVIAAH
   51 RCCNKNRIEE RSQTVKCSCL PGKVAGTTRN RPSCVDASIV IWKWWCEMEP
  101 CLEGEECKTL PDNSGWMCAT GNKIKTTRIH PRT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_4b6, frame 1

```
TREMBLNEW:AF131851_1 product: "Unknown"; Homo sapiens clone 25003 mRNA
sequence, partial cds., N = 1, Score = 242, P = 1.7e-20


>TREMBLNEW:AF131851_1 product: "Unknown"; Homo sapiens clone 25003 mRNA
     sequence, partial cds.
           Length = 165

  HSPs:

 Score = 242 (36.3 bits), Expect = 1.7e-20, P = 1.7e-20
 Identities = 44/89 (49%), Positives = 58/89 (65%)

Query:    42 GTCEVIAAHRCCNKNRIEERSQTVKCSCLPGKVAGTTRNRPSCVDASIVIWKWWCEMEPC 101
             GTCE++    R  ++ R      QT +C+C  G++AGTTR RP+CVDA I+   K WC+M PC
Sbjct:    76 GTCEIVTLDRDSSQPRRTIARQTARCACRKGQIAGTTRARPACVDARIIKTKQWCDMLPC 135

Query:   102 LEGEECKTLPDNSGWMCAT-GNKIKTTRI 129
             LEGE C  L + SGW C   G +IKTT +
Sbjct:   136 LEGEGCDLLINRSGWTCTQPGGRIKTTTV 164
```

Pedant information for DKFZphfkd2_4b6, frame 1

Report for DKFZphfkd2_4b6.1

```
[LENGTH]        133
[MW]            15030.64
[pI]            8.49
[HOMOL]         TREMBLNEW:AF131851_1 product: "Unknown"; Homo sapiens clone 25003 mRNA
sequence, partial cds. 4e-20
[KW]            Alpha_Beta
[KW]            SIGNAL_PEPTIDE 26

SEQ     MAMVSAMSWVLYLWISACAMLLCHGSLQHTFQQHHLHRPEGGTCEVIAAHRCCNKNRIEE
PRD     ccchhhhhhhhhhhhhhhhhhhhcccccchhhhhhhcccccccccceeeeeeecccccchhhh

SEQ     RSQTVKCSCLPGKVAGTTRNRPSCVDASIVIWKWWCEMEPCLEGEECKTLPDNSGWMCAT
PRD     hhhhhhcccccccccccccccccceeeeeehhhhhhcccccccccceeeecccccceeec

SEQ     GNKIKTTRIHPRT
PRD     ccccccccccccc
```

(No Prosite data available for DKFZphfkd2_4b6.1)

(No Pfam data available for DKFZphfkd2_4b6.1)

DKFZphfkd2_4c8
----

group: kidney derived

DKFZphfkd2_4c8 encodes a novel 153 amino acid protein with partial similarity to huntington's associated protein HAP1.

The novel protein contains a leucine zipper involved in protein-protein interaction.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to KIAA0549 and HAP1

potential frame shift at Bp ~1350-1500 will be checked

Sequenced by GBF

Locus: unknown

Insert length: 3182 bp
Poly A stretch at pos. 3162, polyadenylation signal at pos. 3135

```
   1 GGGCTTCCCC CATAGAATTT TTCTTTTCAT TGCCCACTTT ACTGTTTTGG
  51 CTCCAGACTG TCGTTAAGAA TGTACAGCCT AATTCTGGTG TGTTTCGGGA
 101 TATTCTTCTG TCCAGTATTC TGGAAGGGCG GGGAGGCATG GCAGCGTTTT
 151 ACTTGACGTT GATGGTGCTG TGAAGTCCAT TCTTTCCTCT GCAAGACTAC
 201 TGACTATGCA GAAATTTATC GAAGCGGATT ATTATGAACT AGACTGGTAT
 251 TATGAAGAAT GCTCGGATGT TTTATGTGCT GAAAGAGTTG GCCAGATGAC
 301 TAAGACATAT AATGACATAG ATGCTGTCAC TCGGCTTCTT GAGGAGAAAG
 351 AGCGGGATTT AGAATTGGCC GCTCGCATCG GCCAGTCGTT GTTGAAGAAG
 401 AACAAGACCC TAACCGAGAG GAACGAGCTG CTGGAGGAGC AGGTGGAACA
 451 CATCAGGGAG GAGGTGTCTC AGCTCCGGCA TGAGCTGTCC ATGAAGGATG
 501 AGCTGCTTCA GTTCTACACC AGCGCAGCGG AGGAGAGTGA GCCCGAGTCC
 551 GTTTGCTCAA CCCCGTTGAA GAGGAATGAG TCGTCCTCCT CAGTCCAGAA
 601 TTACTTTCAT TTGGATTCTC TTCAAAAGAA GCTGAAAGAC CTTGAAGAGG
 651 AGAATGTTGT ACTTCGATCC GAGGCCAGCC AGCTGAAGAC AGAGACCATC
 701 ACCTATGAGG AGAAGGAGCA GCAGCTGGTC AATGACTGCG TGAAGGAGCT
 751 GAGGGATGCC AATGTCCAGA TTGCTAGTAT CTCAGAGGAA CTGGCCAAGA
 801 AGACGGAAGA TGCTGCCCGC CAGCAAGAGG AGATCACACA CCTGCTATCG
 851 CAAATAGTTG ATTTGCAGAA AAAGGCAAAA GCTTGCGCAG TGGAAAATGA
 901 AGAACTTGTC CAGCATCTGG GGGCTGCTAA GGATGCCCAG CGGCAGCTCA
 951 CAGCCGAGCT GCGTGAGCTG GAGGACAAGT ACGCAGAGTG CATGGAGATG
1001 CTGCATGAGG CGCAGGAGGA GCTGAAGAAC CTCCGGAACA AAACCATGCC
1051 CAATACCACG TCTCGGCGCT ACCACTCACT GGGCCTGTTT CCCATGGATT
1101 CCTTGGCAGC AGAGATTGAG GGAACGATGC GCAAGGAGCT GCAGTTGGAA
1151 GAGGCCGAGT CTCCAGACAT CACTCACCAG AAGCGTGTCT TTGAGACAGT
1201 AAGAAACATC AACCAGGTTG TCAAGCAGAG ATCTCTGACC CCTTCTCCCA
1251 TGAACATCCC CGGCTCCAAC CAGTCCTCGG CCATGAACTC CCTCCTGTCC
1301 AGCTGCGTCA GCACCCCCCG GTCCAGCTTC TACGGCAGCG ACATAGGCAA
1351 CGTCGTCCTC GACAACAAGA CCAACAGCAT CATTCTGGAA ACAGAGGCAG
1401 CCGACCTGGG AAACGATGAG CGGAGTAAGA AGCCGGGGAC GCCGGGCACC
1451 CCCAGGCTCC CACGACCTGG AGACGGCGCT GAGGCGGCTG TCCCTGCGCC
1501 GGGAGAACTA CCTCTCGGAG AGGAGGTTCT TTGAGGAGGA GCAAGAGAGG
1551 AAGCTCCAGG AGCTGGCGGA GAAGGGCGAG CTGCGCAGCG GCTCCCTCAC
1601 ACCCACTGAG AGCATCATGT CCCTGGGCAC GCACTCCCGC TTCTCCGAGT
1651 TCACCGGCTT CTCTGGCATG TCCTTCAGCA GCCGCTCCTA CCTGCCTGAG
1701 AAGCTCCAGA TCGTGAAGCC GCTGGAAGGT GATCACGCGG GGCCTCGGCC
1751 CCTCTCTGTC CTCCTGGGGG ACTCCCTTTG GTCCCTGATC CACCTGCGGA
1801 AGGCGGGGCA CCTCTGTCAC GCCTACTCCT TTTTCTTCCG CGACAGCCAC
1851 CCGCGCTGCT GGTTTGAGTT CCTCTGAGGG TGGTGCTCAG CCTAGGCCTC
1901 CGTCCCTCCC CTCTGGCTGG CAGGTGTGAC AATGCACACA TAGGCCATGA
1951 AACTCGCCGA GGAAAGACAA GCATGTGCAC TGTGGTCTTC TAGTTCTTTC
2001 CTTTGCCTTT AGAACCTTAG AAATAAAAAC TTTTGTGGCG GTAGAGGCAC
2051 TGCTAACTGA TTCAAAAAATT AATTAGGTTT TGCCTGTGGG TGTGAGGAAT
2101 GCAGAAAATT AATGCTTTAG CTTTTCTGCA GTTTTGGTGT CGGGGAGAGG
2151 TTCCAAGCAA ACTCTATTAA ATGGGGATTT TTTTTTCCCC ATAACCACCT
2201 GAATGTGATT TGTGGGCTTA TGTGTTCTGA TTTGAACTTC ATATAGCAAG
2251 GTTGTGGCTT TTGGCAGATG CAGTATGTTC TGAGCGCGGC TCCTAGAGTC
2301 TACAATTTGG AGTCCAGGAA GGGGTGGCTG TGGAGACAAG TGAGTTTTGT
2351 ACCTCCGTAA GCCACCCTTT TTCAGGGTCA GTTCATGTGT TAGTATCAGG
2401 GGCATCTCAG ATGATTAAAC TCATGGGAAA AACTTCCTCC TTCCCTCTCT
2451 CCCTCTTGCC CTCCTGCCTC TTTTTTTTTT TTTTTTTTTT AATTTGGGCA
2501 CTTATAAAAT GTTTTCCCTC TACCTGCTGC TACTCTGCCA AGAGCCACCA
2551 AGTGCTTATA TTTTTCATTT TTTACTCCTT TAGTTTGGAA AGCCATATAC
2601 GTTTGAGAAG GTGTTTTAAA ACTCTGTGTT ACACTTACGA TGCAAAGCCA
2651 AATCAGAACT TCTGTAAGGC AGAACTTTCC CAACTTTAAA AAAATTATTG
```

```
2701 TCCCCTCTAG GAGCCTTCTT AGACGTTTTT TCCTAATCAC CCCCCAAAGA
2751 CATTTTAATA CCACATATAT ATTGTTTATG TACTATATGT ATATACATAA
2801 ACAATACATA AGCAATACAT CTGTGGTATT AAAATTAAAA AGAATCCAAT
2851 TATGTTTACC TCAAAAGAAC CTGTTTTTGC TTCTTGGGAG CAATATTGCC
2901 CCTGTGAGAC TGCATGCTAT AAGGTAAGGT TGTGCTTGTT AAAGACCCAA
2951 GACATGACTG GGTTCCACAG TCTCCAAAGG AAGAGGGTGG GCTAGTTTGT
3001 TTTTATTATT ATTTTAAAAT TGTATAATTG GGGTCTTTCT TAGAGTTCAG
3051 AAAAGGTATA GCTTACTCTT TTTTAATTGT TTATTTAGTT GTAAGCTTAG
3101 TGATTGTTTT CTGATCCACA TTGTGTGTGT TCTTCAATAA AATCTTTCAT
3151 TTCTGCAATT TTAAAAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 206 bp to 1531 bp; peptide length: 442
Category: similarity to known protein
Classification: unset
Prosite motifs: LEUCINE_ZIPPER (139-161)

```
  1 MQKFIEADYY ELDWYYEECS DVLCAERVGQ MTKTYNDIDA VTRLLEEKER
 51 DLELAARIGQ SLLKKNKTLT ERNELLEEQV EHIREEVSQL RHELSMKDEL
101 LQFYTSAAEE SEPESVCSTP LKRNESSSSV QNYFHLDSLQ KKLKDLEEEN
151 VVLRSEASQL KTETITYEEK EQQLVNDCVK ELRDANVQIA SISEELAKKT
201 EDAARQQEEI THLLSQIVDL QKKAKACAVE NEELVQHLGA AKDAQRQLTA
251 ELRELEDKYA ECMEMLHEAQ EELKNLRNKT MPNTTSRRYH SLGLFPMDSL
301 AAEIEGTMRK ELQLEEAESP DITHQKRVFE TVRNINQVVK QRSLTPSPMN
351 IPGSNQSSAM NSLLSSCVST PRSSFYGSDI GNVVLDNKTN SIILETEAAD
401 LGNDERSKKP GTPGTPRLPR PGDGAEAAVP APGELPLGEE VL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_4c8, frame 2

PIR:S72555 huntingtin-associated protein HAP1 - human (fragment), N = 1, Score = 234, P = 8.6e-19

TREMBL:CEUT27A3_7 gene: "T27A3.1"; Caenorhabditis elegans cosmid T27A3., N = 1, Score = 226, P = 9.9e-16

PIR:S67495 huntingtin-associated protein HAP1-A - rat, N = 1, Score = 215, P = 1.6e-14

```
>PIR:S72555 huntingtin-associated protein HAP1 - human (fragment)
          Length = 320

  HSPs:

 Score = 234 (35.1 bits), Expect = 8.6e-19, P = 8.6e-19
 Identities = 66/189 (34%), Positives = 110/189 (58%)

Query:   109 EESEPESVCSTPLKRNE--SSSSVQNYFH---LDSLQKKLKDLEEENVVLRSEASQLKTE 163
             EE+E +  C+ P    +  S  ++ +  H   L++LQ+KL+ LEEEN  LR EASQL T
Sbjct:    28 EEAEEDLQCAHPCDAPKLISQEALLHQHHCPQLEALQEKLRLLEEENHQLREEASQLDT- 86

Query:   164 TITYEEKEQQLVNDCVKELRDANVQIASISEELAKKTEDAARQQEEITHLLSQIVDLQKK 223
                E++EQ L+ +CV++  +A+ Q+A +SE L  + E+  RQQ+E+  L +Q++ LQ++
Sbjct:    87 ---LEDEEQMLILECVEQFSEASQQMAELSEVLVLRLENYERQQQEVARLQAQVLKLQQR 143

Query:   224 AKACAVENEELVQHLGAAKDAQRQLTAE--LRELEDKYAECME--MLHEAQEELKNL-RN 278
               +    E E+L + L + K+ Q QL  E  L   ++  AE +  +    + + + RN
```

```
Sbjct:   144 CRMYGAETEKLQKQLASEKEIQMQLQEEETLPGFQETLAEELRTSLRRMISDPVYFMERN 203

Query:   279 KTMP--NTTSRRY 289
              MP  +T+S RY
Sbjct:   204 YEMPRGDTSSLRY 216
```

Peptide information for frame 3

ORF from 1416 bp to 1874 bp; peptide length: 153
Category: similarity to known protein
Classification: unset

```
  1 MSGVRSRGRR APPGSHDLET ALRRLSLRRE NYLSERRFFE EEQERKLQEL
 51 AEKGELRSGS LTPTESIMSL GTHSRFSEFT GFSGMSFSSR SYLPEKLQIV
101 KPLEGDHAGP RPLSVLLGDS LWSLIHLRKA GHLCHAYSFF FRDSHPRCWF
151 EFL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_4c8, frame 3

TREMBL:AB011121_1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens mRNA for KIAA0549 protein, partial cds., N = 1, Score = 252, P = 5.5e-21

>TREMBL:AB011121_1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens mRNA for KIAA0549 protein, partial cds.
Length = 469

HSPs:

Score = 252 (37.8 bits), Expect = 5.5e-21, P = 5.5e-21
Identities = 57/98 (58%), Positives = 69/98 (70%)

```
Query:     8 GRRAPPGSHDLETALRRLSLRRENYLSERRFFEEEQERKLQELAEKGELRSGSLTPTESI 67
             G+  P G  DL TAL RLSLRR+NYLSE++FF EE +RK+Q LA++ E  SG +TPTES+
Sbjct:    27 GQPGPSGDSDLATALHRLSLRRQNYLSEKQFFAEEWQRKIQVLADQKEGVSGCVTPTESL 86

Query:    68 MSLGTHSRFSEFTGFSGMSFSSRSYLPEKLQIVKPLEG 105
              SL T    SE T  S  S   R ++PEKLQIVKPLEG
Sbjct:    87 ASLCTTQ--SEITDLSSAS-CLRGFMPEKLQIVKPLEG 121
```

Pedant information for DKFZphfkd2_4c8, frame 2

Report for DKFZphfkd2_4c8.2

[LENGTH] 442
[MW] 50020.14
[pI] 4.77
[HOMOL] TREMBL:AF040723_1 product: "neuroan1"; Homo sapiens neuroan1 mRNA, complete cds. 5e-29
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 5e-08
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YIL149c] 5e-08
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-08
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YIL138c] 6e-08
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YGR130c] 2e-07
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-06
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-06
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1643] 1e-06
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-06
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-06
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 4e-06
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 4e-06
[FUNCAT] 03.13 meiosis [S. cerevisiae, YNL250w] 2e-05
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 2e-05

[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c] 5e-05
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 5e-05
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YNL079c] 5e-05
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c] 1e-04
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 1e-04
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YNL272c] 3e-04
[FUNCAT] 08.16 extracellular transport [S. cerevisiae, YNL272c] 3e-04
[BLOCKS] BL01289B
[BLOCKS] BL00415M Synapsins proteins
[EC] 3.6.1.32 Myosin ATPase 2e-07
[PIRKW] tandem repeat 2e-07
[PIRKW] heterodimer 1e-06
[PIRKW] endocytosis 9e-07
[PIRKW] heart 1e-06
[PIRKW] transmembrane protein 4e-07
[PIRKW] zinc finger 9e-07
[PIRKW] metal binding 9e-07
[PIRKW] DNA binding 3e-06
[PIRKW] muscle contraction 2e-07
[PIRKW] acetylated amino end 3e-06
[PIRKW] actin binding 2e-07
[PIRKW] mitosis 1e-06
[PIRKW] microtubule binding 1e-06
[PIRKW] ATP 2e-07
[PIRKW] chromosomal protein 1e-06
[PIRKW] receptor 3e-08
[PIRKW] thick filament 2e-07
[PIRKW] phosphoprotein 8e-06
[PIRKW] glycoprotein 3e-08
[PIRKW] skeletal muscle 3e-06
[PIRKW] DNA condensation 1e-06
[PIRKW] alternative splicing 2e-06
[PIRKW] coiled coil 2e-07
[PIRKW] P-loop 2e-07
[PIRKW] heptad repeat 4e-07
[PIRKW] methylated amino acid 2e-07
[PIRKW] peripheral membrane protein 9e-07
[PIRKW] cardiac muscle 6e-06
[PIRKW] hydrolase 2e-07
[PIRKW] muscle 2e-06
[PIRKW] cytoskeleton 2e-06
[PIRKW] Golgi apparatus 4e-07
[PIRKW] calmodulin binding 9e-07
[SUPFAM] myosin motor domain homology 2e-07
[SUPFAM] tropomyosin TPM1 2e-06
[SUPFAM] giantin 4e-07
[SUPFAM] protein kinase C zinc-binding repeat homology 2e-06
[SUPFAM] human early endosome antigen 1 9e-07
[SUPFAM] unassigned kinesin-related proteins 4e-07
[SUPFAM] M5 protein 8e-08
[SUPFAM] cytoskeletal keratin 3e-06
[SUPFAM] myosin heavy chain 2e-07
[SUPFAM] conserved hypothetical P115 protein 1e-06
[SUPFAM] centromere protein E 1e-06
[SUPFAM] pleckstrin repeat homology 2e-06
[SUPFAM] kinesin motor domain homology 4e-07
[PROSITE] LEUCINE_ZIPPER 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 6.79 %
[KW] COILED_COIL 27.15 %

```
SEQ    MQKFIEADYYELDWYYEECSDVLCAERVGQMTKTYNDIDAVTRLLEEKERDLELAARIGQ
SEG    ..........................................xxxxxxxxxxxxxxx...
PRD    ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  ...........................................................C

SEQ    SLLKKNKTLTERNELLEEQVEHIREEVSQLRHELSMKDELLQFYTSAAEESEPESVCSTP
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC........................

SEQ    LKRNESSSSVQNYFHLDSLQKKLKDLEEENVVLRSEASQLKTETITYEEKEQQLVNDCVK
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  ............CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC...............
```

```
SEQ     ELRDANVQIASISEELAKKTEDAARQQEEITHLLSQIVDLQKKAKACAVENEELVQHLGA
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ..................................................CCCCCCCCCC

SEQ     AKDAQRQLTAELRELEDKYAECMEMLHEAQEELKNLRNKTMPNTTSRRYHSLGLFPMDSL
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....................

SEQ     AAEIEGTMRKELQLEEAESPDITHQKRVFETVRNINQVVKQRSLTPSPMNIPGSNQSSAM
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccchhhh
COILS   ............................................................

SEQ     NSLLSSCVSTPRSSFYGSDIGNVVLDNKTNSIILETEAADLGNDERSKKPGTPGTPRLPR
SEG     .................................................xxxxxxxxxxx
PRD     hhhhhcccccccccccccccccceeeeeccccceeccccccccccccccccccccccccc
COILS   ............................................................

SEQ     PGDGAEAAVPAPGELPLGEEVL
SEG     xxxx..................
PRD     cccccccccccccccccccccc
COILS   ......................
```

Prosite for DKFZphfkd2_4c8.2

PS00029	139->161	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfkd2_4c8.2)

Pedant information for DKFZphfkd2_4c8, frame 3

Report for DKFZphfkd2_4c8.3

[LENGTH] 153
[MW] 17642.03
[pI] 9.38
[HOMOL] TREMBL:AB011121_1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens mRNA for KIAA0549 protein, partial cds. 2e-12
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 12.42 %

```
SEQ     MSGVRSRGRRAPPGSHDLETALRRLSLRRENYLSERRFFEEEQERKLQELAEKGELRSGS
SEG     ..................................xxxxxxxxxxxxxxxxxxx.......
PRD     cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccc

SEQ     LTPTESIMSLGTHSRFSEFTGFSGMSFSSRSYLPEKLQIVKPLEGDHAGPRPLSVLLGDS
SEG     ............................................................
PRD     cccccceeeccccccceeeccccccccccccccchhhhhhhhhcccccccccceeeeeccc

SEQ     LWSLIHLRKAGHLCHAYSFFFRDSHPRCWFEFL
SEG     .................................
PRD     chhhhhhhhccccccceeeeecccccccccccc
```

(No Prosite data available for DKFZphfkd2_4c8.3)

(No Pfam data available for DKFZphfkd2_4c8.3)

DKFZphfkd2_4k14
----

group: intracellular transport and trafficking

DKFZphfkd2_4k14.3 encodes a novel 254 amino acid putative GTP-binding protein nearly identical to Rab6.

Rab proteins are members of the Ras superfamily of GTPases. Rab proteins are localised to the cytoplasmic side of organelles and vesicles involved in the secretory (biosynthetic) and endocytotic pathways in eukaryotic cells. Rab proteins direct the targeting and fusion of transport vesicles to their acceptor membranes.
rab6 is a ubiquitous ras-like GTPase involved in intra-Golgi transport.

The new protein can find application in modulating the transport of vesicles inside the Golgi apparatus.

strong similarity to Rab6

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3084 bp
Poly A stretch at pos. 3061, polyadenylation signal at pos. 3043

```
   1 GGGGCACTCA GCAGGTTGGG CTGCGGCGGC GGCGGCTGGG GAAGCCGAAG
  51 CGCCGCGCGT GAGAGATCCC GGATACATCT GCGGTTTGGG CTCCGCCACC
 101 CTCCGTCTCT CTCCCGCAGG TCTCTGAGCC GGGTGCGGAA GGAGGGAACG
 151 GCCCTAGCCT TGGGAAGCCA AAGCACACCC CTGGCTCCCG CCGACACCGC
 201 CCTCCTTCCC TTCCCAGCCG CGGGCCTCGC TCCGTGCTCG GCTACTCTGC
 251 CGGGAGGCGG CGGCGGCTGC CAGTCTGTGG CGAGCCCTGC TGCCCTCCAG
 301 CCGGGCTTCT CCAGCCGGGC TCCTCCACCG GCCCTTGCAG GGGCACAGAG
 351 AGCTCGGCGC CCGCCCTTCC GCTCGCCTTT TTCGTCAGCC GGCTGGAGGA
 401 GCATCGGTCC GGGAGGTCTC TGGGCTGAGG CGGCGACAGC TCCTCTAGTT
 451 CCACCATGTC CGCGGGCGGA GACTTCGGGA ATCCGCTGAG GAAATTCAAG
 501 CTGGTGTTCC TGGGGGAGCA AAGCGTTGCA AAGACATCTT TGATCACCAG
 551 ATTCAGGTAT GACAGTTTTG ACAACACCTA TCAGGCAATA ATTGGCATTG
 601 ACTTTTTATC AAAAACTATG TACTTGGAGG ATGGAACAAT CGGGCTTCGG
 651 CTGTGGGATA CGGCGGGTCA GGAACGTCTC CGTAGCCTCA TTCCCAGGTA
 701 CATCCGTGAT TCTGCTGCAG CTGTAGTAGT TTACGATATC ACAAATGTTA
 751 ACTCATTCCA GCAAACTACA AAGTGGATTG ATGATGTCAG AACAGAAAGA
 801 GGAAGTGATG TTATCATCAC GCTAGTAGGA AATAGAACAG ATCTTGCTGA
 851 CAAGAGGCAA GTGTCAGTTG AGGAGGGAGA GAGGAAAGCC AAAGGGCTGA
 901 ATGTTACGTT TATTGAAACT AGGGCAAAAA CTGGATACAA TGTAAAGCAG
 951 CTCTTTCGAC GTGTAGCAGC AGCTTTGCCG GGAATGGAAA GCACACAGGA
1001 CGGAAGCAGA GAAGACATGA GTGACATAAA ACTGGAAAAG CCTCAGGAGC
1051 AAACAGTCAG CGAAGGGGGT TGTTCCTGCT ACTCTCCCAT GTCATCTTCA
1101 ACCCTTCCTC AGAAGCCCCC TTACTCTTTC ATTGACTGCA GTGTGAATAT
1151 TGGCTTGAAC CTTTTCCCTT CATTAATAAC GTTTTGCAAT TCATCATTGC
1201 TGCCTGTCTC GTGGAGGTGA TCTATTAGCT TCACAAGCAC AAAAAAAGTC
1251 AGCGTCTTCA TTATTTATAT TTTACAAAAA GCCAAATTAT TTCAGCATAT
1301 TCCGGTGATA ACTTTAAAAA TTAGATACAT TTTCTTAACA TTTTTTTCTT
1351 TTTTAATGTT ATGATAATGT ACTTCAAAAT GATGGAAATC TCAACAGTAT
1401 GAGTATGGCT TGGTTAACGA GCAGTATGTT CACAGCCTGC TTTATCTCTC
1451 CTTGCTCTTC TCACCTCTCC CTTACCCCGT TCCCTATTTC CGTGTTCTTA
1501 CCTAGCCTCC CCCCACTTCC TCAAAACAAA CAAGAGATGG CAAAGCAGCA
1551 GTCCGACCAA GCCCACTGGA ATTATCCTTT AATTTTACAG ATACCACTTG
1601 CTGTAGGCTG TGGACCAAGA TGTCCAGAAT TATTCTTGAG CACTGATGTA
1651 AATTACTTAG ATCTTCTTTG AGGTCAGAAT TCAGCGATCA CGGTAGGCAG
1701 TGCTTGAATG AGAAAAGCCT CCTGGTGCAT CTTCAAAATG AGTCCTAAAG
1751 AACATACTGA GTACTTATAA GTAGCAGAAC ATAAAATGTA TTTCTGACTA
1801 ACACAAATGG TCCTTTCACA TGTGCTTTAT TAGACTCTGG GAGAGAAAAG
1851 TAACCAAGTG CTTCAGAACA GGTTTTTAGT ATTTACTTCT TCATGGTAAG
1901 ATAATGAAGT TCTAATGAAC TATTTCTCCC AAGGTTTTAA AATTGTCAAG
1951 AGTTATTCTG TTTGTTTAAA AAGTAAGAAA CCTCTGTAAG CAATAGATTT
2001 TGCTTGGGTT TTCTTTCTTA AAAAAATAAT ACTATGCAGG CAAGACACCA
2051 TAAAAGTTTA ATTCCTTACA GAAGAACCAG TGGAAGAATT TAAATTTGGC
2101 ACTACGATCA AAACTACTGA ATTAGCAGAA ATAACGATAT CTAAAGCTTA
2151 CCAGCAAAAG AACCCTCAGC AGAATAGCAA AAACTTTGCT CAGGACATTT
2201 GAGGTCAAAT TGAAGACGGA AGACGGAAAC CGGAAACCGT TTTCTTGTAA
2251 GCCCCTAGAG GCAGATCAGG TAAGCATACA TAGTAGAGGG AAAGGAGAGA
2301 ATGGAAATAA AACTGAATAT TATGCAGATT TATGCCTTAT TTTTTAGCAT
2351 TTTTTAAGGT TGGGTCTTTC AGGCTGGTTT TGGTTTGTAT TAGATCTGTA
2401 TAGTTTAGTG ATTTAGTTTT ATATTTAAGC TACGATTAAT ATTTTTTCTT
2451 TGGCGATATT TCTTTGCTTT TTTTTTTTAA CAACTTTCCA TTTTTAGATG
```

```
2501 TTTCGTTGAA TCTATTTAGA GCTTCACCAT GGCAATATGT ATTTCCCTTA
2551 AAACACTGCA AACAAATATA CTAGGAGTGT GCCCTTTTAA TCTTTACTAG
2601 TTATTGTGAG ACTGCTGTGT AAGCTAATAA ACACATTTGT AAAAACATTG
2651 TTTGCAGGAA GAAAACTTCG AGTTACAGGT CAGGAAAAGC CTGCTGAATT
2701 TATGTTGTAA ACGTTACTTA ACACAGTATA AAGATGAAAA GACAACAAAA
2751 GTATCTTCAT ACTTCCTCAT CCCCTCATTG CAACAAAACC TTAAACTGGG
2801 AGAACCTTAG TCCCCTCTCT TTCCTCTTCC TCCTCCACTT CCCACTTATT
2851 GCCACTTTGT AATATTCAGA GAGCACTTGG ATTATGGATC TGAATAGAGA
2901 AATGCTTACA GATAATCATT AGCCCACATA CCAGTAACTT ATACTTAAAG
2951 ATGGGATGGA GTTATAAAGT GCTTTTATAA TCCAATATAA TTGCTAAAGG
3001 CAAGGGTTGA CTCTTTGTTT TATTTTGACA TGGCATGTCC TGAAATAAAT
3051 ATTGGTTCAC TATGAAAAAA AAAAAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

98382468:
Rab proteins.

97203146:
GTP-bound forms of rab6 induce the redistribution of Golgi proteins into the endoplasmic reticulum.

Peptide information for frame 3

ORF from 456 bp to 1217 bp; peptide length: 254
Category: strong similarity to known protein
Classification: unset
Prosite motifs: BACTERIAL_OPSIN_RET (45-57)

```
  1 MSAGGDFGNP LRKFKLVFLG EQSVAKTSLI TRFRYDSFDN TYQAIIGIDF
 51 LSKTMYLEDG TIGLRLWDTA GQERLRSLIP RYIRDSAAAV VVYDITNVNS
101 FQQTTKWIDD VRTERGSDVI ITLVGNRTDL ADKRQVSVEE GERKAKGLNV
151 TFIETRAKTG YNVKQLFRRV AAALPGMEST QDGSREDMSD IKLEKPQEQT
201 VSEGGCSCYS PMSSSTLPQK PPYSFIDCSV NIGLNLFPSL ITFCNSSLLP
251 VSWR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_4k14, frame 3

PIR:G34323 GTP-binding protein Rab6 - human, N = 1, Score = 944, P = 6.5e-95

TREMBL:CET25G12_2 gene: "T25G12.4"; Caenorhabditis elegans cosmid T25G12., N = 1, Score = 756, P = 5.4e-75

TREMBL:NTNTRAF_1 gene: "Nt-rab6"; Nicotiana tabacum SR1 Nt-rab6 mRNA, complete cds., N = 1, Score = 698, P = 7.6e-69

TREMBL:D84314_1 product: "rab6"; Drosophila melanogaster mRNA for rab6, complete cds., N = 1, Score = 836, P = 1.9e-83

PIR:T01588 small GTP-binding protein F16B22.10 - Arabidopsis thaliana, N = 1, Score = 704, P = 1.8e-69

>PIR:G34323 GTP-binding protein Rab6 - human
Length = 208

HSPs:

Score = 944 (141.6 bits), Expect = 6.5e-95, P = 6.5e-95
Identities = 186/208 (89%), Positives = 190/208 (91%)

```
Query:    1 MSAGGDFGNPLRKFKLVFLGEQSVAKTSLITRFRYDSFDNTYQAIIGIDFLSKTMYLEDG 60
            MS GGDFGNPLRKFKLVFLGEQSV KTSLITRF YDSFDNTYQA IGIDFLSKTMYLED
Sbjct:    1 MSTGGDFGNPLRKFKLVFLGEQSVGKTSLITRFMYDSFDNTYQATIGIDFLSKTMYLEDR 60

Query:   61 TIGLRLWDTAGQERLRSLIPRYIRDSAAAVVVYDITNVNSFQQTTKWIDDVRTERGSDVI 120
            T+ L+LWDTAGQER RSLIP YIRDS  AVVVYDITNVNSFQQTTKWIDDVRTERGSDVI
Sbjct:   61 TVRLQLWDTAGQERFRSLIPSYIRDSTVAVVVYDITNVNSFQQTTKWIDDVRTERGSDVI 120

Query:  121 ITLVGNRTDLADKRQVSVEEGERKAKGLNVTFIETRAKTGYNVKQLFRRVAAALPGMEST 180
            I LVGN+TDLADKRQVS+EEGERKAK LNV FIET AK GYNVKQLFRRVAAALPGMEST
Sbjct:  121 IMLVGNKTDLADKRQVSIEEGERKAKELNVMFIETSAKAGYNVKQLFRRVAAALPGMEST 180

Query:  181 QDGSREDMSDIKLEKPQEQTVSEGGCSC 208
            QD SREDM DIKLEKPQEQ VSEGGCSC
Sbjct:  181 QDRSREDMIDIKLEKPQEQPVSEGGCSC 208
```

Pedant information for DKFZphfkd2_4k14, frame 3

Report for DKFZphfkd2_4k14.3

```
[LENGTH]     254
[MW]         28385.29
[pI]         7.58
[HOMOL]      PIR:G34323 GTP-binding protein Rab6 - human 1e-102
[FUNCAT]     08.07 vesicular transport (golgi network, etc.)    [S. cerevisiae, YLR262c] 7e-60
[FUNCAT]     30.08 organization of golgi   [S. cerevisiae, YLR262c] 7e-60
[FUNCAT]     30.09 organization of intracellular transport vesicles    [S. cerevisiae, YOR089c] 2e-33
[FUNCAT]     08.19 cellular import [S. cerevisiae, YOR089c] 2e-33
[FUNCAT]     08.13 vacuolar transport    [S. cerevisiae, YOR089c] 2e-33
[FUNCAT]     06.04 protein targeting, sorting and translocation  [S. cerevisiae, YOR089c] 2e-33
[FUNCAT]     09.09 biogenesis of intracellular transport vesicles     [S. cerevisiae, YGL210w] 3e-28
[FUNCAT]     30.02 organization of plasma membrane      [S. cerevisiae, YFL005w] 8e-27
[FUNCAT]     03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w] 8e-27
[FUNCAT]     01.05.04 regulation of carbohydrate utilization     [S. cerevisiae, YOR101w] 2e-21
[FUNCAT]     11.10 cell death      [S. cerevisiae, YOR101w] 2e-21
[FUNCAT]     01.03.13 regulation of nucleotide metabolism      [S. cerevisiae, YOR101w] 2e-21
[FUNCAT]     30.03 organization of cytoplasm      [S. cerevisiae, YOR101w] 2e-21
[FUNCAT]     03.99 other cell growth, cell division and dna synthesis activities      [S. cerevisiae, YOR101w] 2e-21
[FUNCAT]     10.04.07 g-proteins   [S. cerevisiae, YOR101w] 2e-21
[FUNCAT]     03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 6e-19
[FUNCAT]     11.01 stress response [S. cerevisiae, YNL098c] 6e-19
[FUNCAT]     03.10 sporulation and germination    [S. cerevisiae, YNL098c] 6e-19
[FUNCAT]     04.07 rna transport    [S. cerevisiae, YOR185c] 6e-16
[FUNCAT]     30.10 nuclear organization     [S. cerevisiae, YOR185c] 6e-16
[FUNCAT]     08.01 nuclear transport        [S. cerevisiae, YOR185c] 6e-16
[FUNCAT]     30.04 organization of cytoskeleton    [S. cerevisiae, YPR165w] 4e-13
[FUNCAT]     10.02.07 g-proteins    [S. cerevisiae, YPR165w] 4e-13
[FUNCAT]     10.99 other signal-transduction activities   [S. cerevisiae, YCR027c] 2e-09
[FUNCAT]     10.05.07 g-proteins    [S. cerevisiae, YLR229c] 8e-08
[FUNCAT]     03.07 pheromone response, mating-type determination, sex-specific proteins
      [S. cerevisiae, YLR229c] 8e-08
[FUNCAT]     03.01 cell growth      [S. cerevisiae, YNL180c] 1e-05
[FUNCAT]     06.10 assembly of protein complexes  [S. cerevisiae, YOR094w] 5e-05
[BLOCKS]     BL01115A GTP-binding nuclear protein ran proteins
[SCOP]       d1as3_2 3.29.1.4.12 Transducin (alpha subunit), insertion domai 1e-32
[SCOP]       d1mh1__ 3.29.1.4.2 Rac1  [Human (Homo sapiens)  2e-51
[SCOP]       d5p21__ 3.29.1.4.1 cH-p21 Ras protein  [human (Homo sapiens)  7e-53
[SCOP]       d1hura_ 3.29.1.4.8 ADP-ribosylation factor 1 (ARF1)  [human (Hom 1e-46
[SCOP]       d1a2kc_ 3.29.1.4.5 Ran Nuclear transport factor-2 (NTF2)  [Do 6e-60
[PIRKW]      nucleus 2e-14
[PIRKW]      cell cycle control 5e-15
[PIRKW]      membrane trafficking 3e-71
[PIRKW]      endoplasmic reticulum 1e-29
[PIRKW]      phosphoprotein 1e-29
[PIRKW]      prenylated cysteine 2e-36
[PIRKW]      signal transduction 5e-15
[PIRKW]      transforming protein 5e-30
[PIRKW]      purine nucleotide binding 1e-28
[PIRKW]      alternative splicing 1e-18
[PIRKW]      P-loop 3e-71
```

```
[PIRKW]         lipoprotein 2e-36
[PIRKW]         proto-oncogene 1e-20
[PIRKW]         methylated carboxyl end 1e-20
[PIRKW]         membrane protein 1e-29
[PIRKW]         GTP binding 3e-71
[PIRKW]         thiolester bond 1e-29
[PIRKW]         Golgi apparatus 1e-29
[SUPFAM]        ras transforming protein 1e-76
[PROSITE]       BACTERIAL_OPSIN_RET    1
[PFAM]          Ras family (contains ATP/GTP binding P-loop)
[KW]            Alpha_Beta
[KW]            3D

SEQ     MSAGGDFGNPLRKFKLVFLGEQSVAKTSLITRFRYDSFDNTYQAIIGIDFLSKTMYLEDG
1kao-   ..........CCEEEEEEEECTTTTCHHHHHHHHHHCCCCCCCTTTTC-EEEEEEEEETTE

SEQ     TIGLRLWDTAGQERLRSLIPRYIRDSAAAVVVYDITNVNSFQQTTKWIDDVRTERGSDVI
1kao-   EEEEEEEECCTTTTCHHHHHHHHHHHCCEEEEEEETTTHHHHHHHHHHHHHHHHHTTTCCC

SEQ     ITLVGNRTDLADKRQVSVEEGERKAKGLNVTFIETRAKTGYNVKQLFRRVAAALPGMEST
1kao-   EEEEEETTTTGGGCCCCHHHHHHHHHHHCCCEEECTTTTHHHHHHHHHHH..........

SEQ     QDGSREDMSDIKLEKPQEQTVSEGGCSCYSPMSSSTLPQKPPYSFIDCSVNIGLNLFPSL
1kao-   ............................................................

SEQ     ITFCNSSLLPVSWR
1kao-   ..............
```

Prosite for DKFZphfkd2_4k14.3

```
PS00327      45->57    BACTERIAL_OPSIN_RET     PDOC00291
```

Pfam for DKFZphfkd2_4k14.3

```
HMM_NAME       Ras family (contains ATP/GTP binding P-loop)

HMM                *KLVLIGDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGKtIK
                    KLV++G+ +V K++L  RF +++F++ Y + IG+DF++KT+++++ TI
Query          15   KLVFLGEQSVAKTSLITRFRYDSFDNTYQAIIGIDFLSKTMYLEDGTIG      63

HMM                LQIWDTAGQERYRsMRPMYYRGAMGFMLVYDITNRqSFENIrNWweEIrR
                   L +WDTAGQER RS+ P Y+R++ ++++VYDITN  SF+  ++W++++R+
Query          64  LRLWDTAGQERLRSLIPRYIRDSAAAVVVYDITNVNSFQQTTKWIDDVRT    113

HMM                HCDrDENVPIMLVGNKCDLEDQRQVStEEGQeFAREWGAIPFMETSAKTN
                     +  ++V+I LVGN +DL+D+RQVS EEG+  A+ ++ + F+ET AKT+
Query         114  ERG--SDVIITLVGNRTDLADKRQVSVEEGERKAKGLN-VTFIETRAKTG    160

HMM                iNVEEAFMEIvReIlqrMqe.q.NqteNinidQpsrnrk....rCCCIM*
                   +NV++ F +++ +++ +++  + +++++++I+ ++++ +     +C+ +
Query         161  YNVKQLFRRVAAALPGMESTQDGSREDMSDIKLEKPQEQTVSEGGCS-C    208
```

DKFZphfkd2_4m11
----

group: transmembrane protein

DKFZphfbr2-4m11 encodes a novel 159 amino acid protein with weak similarity to the putative membrane protein YMR034c of S. cerevisiae.

The novel protein contains 4 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker of neuronal cells.

weak similarity to YMR034c

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1749 bp
Poly A stretch at pos. 1727, polyadenylation signal at pos. 1713

```
   1 GGGGTCCTCA AAGCCGCCGG AGCAACCCCC AGGTCTTTAC TTTACAATCG
  51 GCAATTTGAC TTGCTCTGCT GCATGTCTGG AGGGACCAAG GAAAGTGTGG
 101 AGACGCTCCA AGGATTAGGT GATCGGAGCT TGAAAAGAAA AAAAGCCAAA
 151 CAAATAAACA AAACCCACCC ACCCTAACGA ATATGAGGCT GCTGGAGAGA
 201 ATGAGGAAAG ACTGGTTCAT GGTCGGAATA GTGCTGGCGA TCGCTGGAGC
 251 TAAACTGGAG CCGTCCATAG GGGTGAATGG GGGACCACTG AAGCCAGAAA
 301 TAACTGTATC CTACATTGCT GTTGCAACAA TATTCTTTAA CAGTGGACTA
 351 TCATTGAAAA CAGAGGAGCT GACCAGTGCT TTGGTGCATC TAAAACTGCA
 401 TCTTTTTATT CAGATCTTTA CTCTTGCATT CTTCCCAGCA ACAATATGGC
 451 TTTTTCTTCA GCTTTTATCA ATCACACCCA TCAACGAATG GCTTTTAAAA
 501 GGTTTGCAGA CAGTAGGTTG CATGCCTCCG CCTGTGTCTT CTGCAGTGAT
 551 TTTAACCAAG GCAGTTGGTG GAAATGAGGC AGCTGCAATA TTTAATTCAG
 601 CCTTTGGAAG TTTTTTGGTA AGTAAACATA GTTTAACTTG TCTATTACAA
 651 CTTTTGCTGT GATATTGTGT ATATGAAAGA TTTAGTGAAA GCTGGATTTG
 701 TTTTACTCTT TGGTTAAGTA TAAAAATTGT TGAATCTTTT CATGTGCCAG
 751 TATCCATACC CTGAAGAAAA GTAGTTAATG AATAAAGCAA ATGTTCTCTT
 801 ACAATATATT TTGGAGGTTT GGATTTTAAA ATTCCATTTA ATGAATTCAA
 851 GGAATCAATT AAAACACTAT GTGTCTCCTT ATAGAGGTTA TGTCAATATA
 901 TTGATCATTT AATGAGGTCT TTTAGATTAT TATTATTTTG TATCATGGGA
 951 CTGAGGATTT TGAAAAGGAA ACATGACCCA GCTGGTCAGA AAGGGAATGC
1001 TAATTTACTT GTTGACATGC CATTTATTTT GTACATTTCA CTGTCAAAGA
1051 AGCTACTGGC TTGGATGCTT CTGAGAAATC TATGTGAGAA AAAATTTGAA
1101 AGGAAGATAT GACTAATGAG TAATTTGCAA GTAAATGTTG TATCTATATA
1151 TATATATATA TAAAGATTCA AAAGTAGTTC AGCTTTCATA AGTAGAACCA
1201 ATATAAGGAC GTTGTTTTAG CATTTTTAAT CATTATTTTT AAATAAATGA
1251 TGTAACAGAG GCTTGATTTG TGTTATGAAA GATTGAGAAA CTAAATTTTC
1301 TGTTGATTTA ATTTTTTTGT GCCTTAAAAC TTTGTTAAAT TCCTGAAGTT
1351 AATTATCATA TTGTACTTTT TGGGGCATAA CTCATTAGCA GATATGTAGT
1401 GCAGTGATTT ACAAATAATT GAGAGTAAAA TCAGTGATGT ATAAACTAGT
1451 TCATGAGTCT AGGTAAAATA TCAATTACCT CTGTTTAAAA TGCTCTGTTA
1501 ATTATTATTG TATGTATTTA AATGTAGTTA AAGCTTTTAA ACATGTTGTT
1551 ACATAGTGTT AATTCTACAC AGTGCTACAC AGCTTTTAGT GTCACATAGC
1601 CTTACAGAGT TTATAATGAT GTAGCATCTG CAAAATATAT GCATAGCTTA
1651 TATCCTATTT TTATAGAGCC AGTAATGGTT TTTGTGATGC TGTATTACTT
1701 CTGGGTTTTA GACAATAAAG TCTGTTTAAC AAAAAAAAAA AAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

```
ORF from 183 bp to 659 bp; peptide length: 159
Category: similarity to unknown protein


    1 MRLLERMRKD WFMVGIVLAI AGAKLEPSIG VNGGPLKPEI TVSYIAVATI
   51 FFNSGLSLKT EELTSALVHL KLHLFIQIFT LAFFPATIWL FLQLLSITPI
  101 NEWLLKGLQT VGCMPPPVSS AVILTKAVGG NEAAAIFNSA FGSFLVSKHS
  151 LTCLLQLLL

                             BLASTP hits

No BLASTP hits available

            Alert BLASTP hits for DKFZphfkd2_4m11, frame 3

PIR:S53951 probable membrane protein YMR034c - yeast (Saccharomyces
cerevisiae), N = 1, Score = 171, P = 3.2e-12

PIR:A65015 yfeH protein - Escherichia coli (strain K-12), N = 1, Score
= 131, P = 4.2e-08


>PIR:S53951 probable membrane protein YMR034c - yeast (Saccharomyces
     cerevisiae)
            Length = 434

  HSPs:

 Score = 171 (25.7 bits), Expect = 3.2e-12, P = 3.2e-12
 Identities = 38/144 (26%), Positives = 72/144 (50%)

Query:      5 ERMRKDWFMVGIVLAIAGAKLEPSIGVNGGPLKPEITVSYIAVATIFFNSGLSLKTEELT 64
              E ++  WF + + + I   A+  P+    +GG +K + ++ Y   VA IF   SGL +K+  L
Sbjct:     18 EFLKSQWFFICLAILIVIARFAPNFARDGGLIKGQYSIGYGCVAWIFLQSGLGMKSRSLM 77

Query:     65 SALVHLKLHLFIQIFTLAFFPATIWLF---LQLLSITPINEWLLKGLQTVGCMPPPVSSA 121
              + +++ + H  I + +      + ++ F    ++  +   I++W+L GL        P  V+S
Sbjct:     78 ANMLNWRAHATILVLSFLITSSIVYGFCCAVKAANDPKIDDWVLIGLILTATCPTTVASN 137

Query:    122 VILTKAVGGNEAAAIFNSAFGSFL 145
              VI+T    GGN      +       G+ L
Sbjct:    138 VIMTTNAGGNSLLCVCEVFIGNLL 161


            Pedant information for DKFZphfkd2_4m11, frame 3
            -----------------------------------------------

                   Report for DKFZphfkd2_4m11.3

[LENGTH]      159
[MW]          17282.92
[pI]          9.06
[HOMOL]       PIR:S53951 probable membrane protein YMR034c - yeast (Saccharomyces cerevisiae)
5e-12
[FUNCAT]      99 unclassified proteins        [S. cerevisiae, YMR034c] 2e-13
[PROSITE]     MYRISTYL          2
[PROSITE]     PKC_PHOSPHO_SITE        1
[KW]          TRANSMEMBRANE   4


SEQ    MRLLERMRKDWFMVGIVLAIAGAKLEPSIGVNGGPLKPEITVSYIAVATIFFNSGLSLKT
PRD    ccchhhhhhhhhhhhhhhhhhhhhhccccccccccccccceeeeeeeccccccccccchhhh
MEM    .....MMMMMMMMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMM..

SEQ    EELTSALVHLKLHLFIQIFTLAFFPATIWLFLQLLSITPINEWLLKGLQTVGCMPPPVSS
PRD    hhhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhccccchhhhhhhheeeeccccccc
MEM    .......MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM......................

SEQ    AVILTKAVGGNEAAAIFNSAFGSFLVSKHSLTCLLQLLL
PRD    ceeeeecccccchhhhhhhccccceeecceeeeeeeccc
MEM    ...........MMMMMMMMMMMMMMMMMMMMMMMMMMMM


                  Prosite for DKFZphfkd2_4m11.3

PS00005        57->60   PKC_PHOSPHO_SITE       PDOC00005
PS00008        15->21   MYRISTYL               PDOC00008
PS00008       129->135  MYRISTYL               PDOC00008

(No Pfam data available for DKFZphfkd2_4m11.3)
```

PAGE INTENTIONALLY LEFT BLANK

DKFZphutel_17k7
----

group: uterus derived

DKFZphutel_17k7 encodes a novel 520 amino acid protein with weak similarity to S. Cerevisiae Fipl.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to S.cerevisiae Fipl

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1914 bp
Poly A stretch at pos. 1897, polyadenylation signal at pos. 1867

```
   1 CGGACGCGTG GGCGGACGCG TGGGGCCTTC CTGGGATTGG AGTCTCGAGC
  51 TTTCTTCGTT CGTTCGCCGG CGGGTTCGCG CCCTTCTCGC GCCTCGGGGC
 101 TGCGAGGCTG GGGAAGGGGT TGGAGGGGGC TGTTGATCGC CGCGTTTAAG
 151 TTGCGCTCGG GGCGGCCATG TCGGCCGGCG AGGTCGAGCG CCTAGTGTCG
 201 GAGCTGAGCG GCGGGACCGG AGGGGATGAG GAGGAAGAGT GGCTCTATGG
 251 CGATGAAAAT GAAGTTGAAA GGCCAGAAGA AGAAAATGCC AGTGCTAATC
 301 CTCCATCTGG AATTGAAGAT GAAACTGCTG AAAATGGTGT ACCAAAACCG
 351 AAAGTGACTG AGACCGAAGA TGATAGTGAT AGTGACAGCG ATGATGATGA
 401 AGATGATGTT CATGTCACTA TAGGAGACAT TAAAACGGGA GCACCACAGT
 451 ATGGGAGTTA TGGTACAGCA CCTGTAAATC TTAACATCAA GACAGGGGGA
 501 AGAGTTTATG GAACTACAGG GACAAAAGTC AAAGGAGTAG ACCTTGATGC
 551 ACCTGGAAGC ATTAATGGAG TTCCACTCTT AGAGGTAGAT TTGGATTCTT
 601 TTGAAGATAA ACCATGGCGT AAACCTGGTG CTGATCTTTC TGATTATTTT
 651 AATTATGGGT TTAATGAAGA TACCTGGAAA GCTTACTGTG AAAAACAAAA
 701 GAGGATACGA ATGGGACTTG AAGTTATACC AGTAACCTCT ACTACAAATA
 751 AAATTACGGT ACAGCAGGGA AGAACTGGAA ACTCAGAGAA AGAAACTGCC
 801 CTTCCATCTA CAAAAGCTGA GTTTACTTCT CCTCCTTCTT TGTTCAAGAC
 851 TGGGCTTCCA CCGAGCAGGA GATTACCTGG GGCAATTGAT GTTATCGGTC
 901 AGACTATAAC TATCAGCCGA GTAGAAGGCA GGCGACGGGC AAATGAGAAC
 951 AGCAACATAC AGGTCCTTTC TGAAAGATCT GCTACTGAAG TAGACAACAA
1001 TTTTAGCAAA CCACCTCCGT TTTTCCCTCC AGGAGCTCCT CCCACTCACC
1051 TTCCACCTCC TCCATTTCTT CCACCTCCTC CGACTGTCAG CACTGCTCCA
1101 CCTCTGATTC CACCACCGGG TTTTCCTCCT CCACCAGGCG CTCCACCTCC
1151 ATCTCTTATA CCAACAATAG AAAGTGGACA TTCCTCTGGT TATGATAGTC
1201 GTTCTGCACG TGCATTTCCA TATGGCAATG TTGCCTTTCC CCATCTTCCT
1251 GGTTCTGCTC CTTCGTGGCC TAGTCTTGTG GACACCAGCA AGCAGTGGGA
1301 CTATTATGCC AGAAGAGAGA AAGACCGAGA TAGAGAGAGA GACAGAGACA
1351 GAGAGCGAGA CCGTGATCGG GACAGAGAAA GAGAACGCAC CAGAGAGAGA
1401 GAGAGGGAGC GTGATCACAG TCCTACACCA AGTGTTTTCA ACAGCGATGA
1451 AGAACGATAC AGATACAGGG AATATGCAGA AAGAGGTTAT GAGCGTCACA
1501 GAGCAAGTCG AGAAAAAGAA GAACGACATA GAGAAAGACG ACACAGGGAG
1551 AAAGAGGAAA CCAGACATAA GTCTTCTCGA AGTAATAGTA GACGTCGCCA
1601 TGAAAGTGAA GAAGGAGATA GTCACAGGAG ACACAAACAC AAAAAATCTA
1651 AAAGAAGCAA AGAAGGAAAA GAAGCGGGCA GTGAGCCTGC CCCTGAACAG
1701 GAGAGCACCG AAGCTACACC TGCAGAATAG GCATGGTTTT GGCCTTTTGT
1751 GTATATTAGT ACCAGAAGTA GATACTATAA ATCTTGTTAT TTTTCTGGAT
1801 AATGTTTAAG AAATTTACCT TAAATCTTGT TCTGTTTGTT AGTATGAAAA
1851 GTTAACTTTT TTTCCAAAAT AAAAGAGTGA ATTTTTCATG TTAAGTTAAA
1901 AAAAAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 168 bp to 1727 bp; peptide length: 520
Category: similarity to known protein

```
  1 MSAGEVERLV SELSGGTGGD EEEEWLYGDE NEVERPEEEN ASANPPSGIE
 51 DETAENGVPK PKVTETEDDS DSDSDDDEDD VHVTIGDIKT GAPQYGSYGT
101 APVNLNIKTG GRVYGTTGTK VKGVDLDAPG SINGVPLLEV DLDSFEDKPW
151 RKPGADLSDY FNYGFNEDTW KAYCEKQKRI RMGLEVIPVT STTNKITVQQ
201 GRTGNSEKET ALPSTKAEFT SPPSLFKTGL PPSRRLPGAI DVIGQTITIS
251 RVEGRRRANE NSNIQVLSER SATEVDNNFS KPPPFFPPGA PPTHLPPPPF
301 LPPPPTVSTA PPLIPPPGFP PPPGAPPPSL IPTIESGHSS GYDSRSARAF
351 PYGNVAFPHL PGSAPSWPSL VDTSKQWDYY ARREKDRDRE RDRDRERDRD
401 RDRERERTRE RERERDHSPT PSVFNSDEER YRYREYAERG YERHRASREK
451 EERHRERRHR EKEETRHKSS RSNSRRRHES EEGDSHRRHK HKKSKRSKEG
501 KEAGSEPAPE QESTEATPAE
```

BLASTP hits

Entry AF016427_4 from database TREMBL:
gene: "F32D1.9"; Caenorhabditis elegans cosmid F32D1.
Score = 392, P = 1.8e-36, identities = 156/519, positives = 212/519

Entry S62454 from database PIR:
hypothetical protein SPAC22G7.10 - fission yeast (Schizosaccharomyces pombe)
Score = 246, P = 2.0e-22, identities = 62/163, positives = 91/163

Entry A56545 from database PIR:
FIP1 protein - yeast (Saccharomyces cerevisiae)
Score = 186, P = 2.9e-16, identities = 56/206, positives = 92/206

Alert BLASTP hits for DKFZphutel_17k7, frame 3

TREMBLNEW:AF109907_1 product: "S164"; Homo sapiens S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds; and S171 gene, partial cds., N = 2, Score = 236, P = 1.5e-16

```
>TREMBLNEW:AF109907_1 product: "S164"; Homo sapiens S164 gene, partial cds;
     PS1 and hypothetical protein genes, complete cds; and S171 gene, partial
     cds.
            Length = 735

  HSPs:

 Score = 236 (35.4 bits), Expect = 1.5e-16, Sum P(2) = 1.5e-16
 Identities = 51/120 (42%), Positives = 76/120 (63%)

Query:   383 REKDRDRERDRDRERDRDRDRERERTRERERERDHSPTPSVFNSDEERYRYREYA---ER 439
             REK+++RER+R+R+RDRDR +ER+R R+RER+RD     S  + +++R R RE +   ER
Sbjct:   227 REKEKERERERERDRDRDRTKERDRDRDRERDRDRDRERSS-DRNKDRSRSREKSRDRER 285

Query:   440 GYERHRASREKEERHRER-RHREKEETRHKSSRSNSRRRHESEEGDSHRRHKHKKSKRSK 498
               ER R     + ER RER R RE+E  R +     + +R  E +E D++ R K ++  R K
Sbjct:   286 ERERERERERERERERERERERERERERERERKDKKRDREEDEEDAYERRKLERKLREK 345

Query:   499 E 499
             E
Sbjct:   346 E 346

 Score = 214 (32.1 bits), Expect = 4.4e-14, Sum P(2) = 4.4e-14
 Identities = 50/133 (37%), Positives = 75/133 (56%)

Query:   383 REKDRDR-ERDRDRERDRDRDRERERTRERERERDHSPTPSVFNS-DEERYRYREYAERG 440
             RE++R+R ER+R+RER+R+R++E+ER RERER+RD   T        D ER R R+  ER
Sbjct:   208 RERERRERERERERERERERERKEKEREREREDRDRDRTKERDRDRDRERDRDRD-RERS 266

Query:   441 YERHRASREKEERHRERRHREKEETRHKSSRSNSRRRHESEEGDSHRRHKHKKSKRSKEG 500
              +R++        E+ R+R  RE+E  R +  R   R R     E +  R  + ++ K  K
Sbjct:   267 SDRNKDRSRSREKSRDRE-RERERERE-RERERERERERERERERERERERERERKDKKRD 324

Query:   501 KEAGSEPAPEQESTE 515
             +E   E A E+   E
Sbjct:   325 REEDEEDAYERRKLE 339
```

```
 Score = 214 (32.1 bits), Expect = 4.4e-14, Sum P(2) = 4.4e-14
 Identities = 55/141 (39%), Positives = 80/141 (56%)

Query:    383 REKDRDR-ERDRDRERDRDRDRERERTRERERERDHSPTPSVFNS-DEERYRYREYAERG 440
              RE++R+R ER+R+RER+R+R++E+ER RERER+RD   T       D ER R R+  ER
Sbjct:   208 RERERERRERERERERERERERKEKERERERERDRDRDRTKERDRDRDRERDRDRD-RERS 266

Query:    441 YERHR-ASREKEE-RHRER-RHREKEETRHKSSRSNSRRRHESEEGDSHRRHKHKKSKRS 497
               +R++  SR +E+ R RER R RE+E  R +         R  E E       R K KK  R
Sbjct:   267 SDRNKDRSRSREKSRDREREREREREREREREREREREREREREREREREREKDKKRDRE 326

Query:    498 KEGKEAGSEPAPEQESTEATPA 519
              ++ ++A      E++  E   A
Sbjct:   327 EDEEDAYERRKLERKLREKEAA 348

 Score = 210 (31.5 bits), Expect = 1.2e-13, Sum P(2) = 1.2e-13
 Identities = 59/142 (41%), Positives = 78/142 (54%)

Query:    383 REKDRDRERDRDRERDRDRDRERERTRERERERDHSPTPSVFNS---DEERYRYREYAER 439
              RE++RDR+RDR +ERDRDRDRER+R R+RER  D +   S         D ER R RE  ER
Sbjct:   235 RERERDRDRDRTKERDRDRDRERDRDRDRERSSDRNKDRSRSREKSRDRERERERE-RER 293

Query:    440 GYERHRA-SREKE-ERHRER-RHREKEETRHKSS-------RSNSRRRHESEEGDSHRRH 489
                ER R   RE+E ER RER R REK++ R +           R    R+  +E      R
Sbjct:   294 ERERERERERERERERERERERERERERERKDKKRDREEDEEDAYERRKLERKLREKEAAYQERL 353

Query:    490 KHKKSKRSKEGKEAGSEPAPEQE 512
              K+ + +  K+ +E   E  E+E
Sbjct:   354 KNWEIRERKKTREYEKEAEREEE 376

 Score = 205 (30.8 bits), Expect = 4.4e-13, Sum P(2) = 4.4e-13
 Identities = 59/149 (39%), Positives = 83/149 (55%)

Query:    372 DTSKQWDYYARREKDRDR--ERDRDRERDRDRDRERERTRERERERDHSPTPSVFNSDEE 429
              +  K+ +     R++DRDR  ERDRDR+R+RDRDR+RER+ +R ++R  S   S    D E
Sbjct:   228 EKEKERERERERDRDRDRTKERDRDRDRERDRDRDRERSSDRNKDRSRSREKS---RDRE 284

Query:    430 RYRYREYAERGYERHRA-SREKE-ERHRER-RHREKEETRHKSS-------RSNSRRRHE 479
              R R RE  ER  ER R   RE+E ER RER R REK++ R +          R    R+
Sbjct:   285 RERE-RERERERERERERERERERERERERERERKDKKRDREEDEEDAYERRKLERKLR 343

Query:    480 SEEGDSHRRHKHKKSKRSKEGKEAGSEPAPEQE 512
               +E     R K+ + +  K+ +E   E  E+E
Sbjct:   344 EKEAAYQERLKNWEIRERKKTREYEKEAEREEE 376

 Score = 202 (30.3 bits), Expect = 9.6e-13, Sum P(2) = 9.6e-13
 Identities = 49/117 (41%), Positives = 70/117 (59%)

Query:    383 REKDRDRERDRDRERDRDRDRERERTRERERERDHSPTPSVFNSDEERYRYREYAERGYE 442
              REK RDRER+R+RER+R+R+RERER RERERER+           D++R R  E  E  YE
Sbjct:   277 REKSRDRERERERERERERERERERERERERERER-EKDKKRDR-EEDEEDAYE 334

Query:    443 RHRASREKEERHRERRHREKEETRHKSSRSNSRR-RHESEEGDSHRRHKHKKSKRSKE 499
              R +   E++ R +E  ++E+ +        R  +R    E+E  +  RR   K++KR KE
Sbjct:   335 RRKL--ERKLREKEAAYQERLKNWEIRERKKTREYEKEAEREEERRREMAKEAKRLKE 390

 Score = 183 (27.5 bits), Expect = 1.2e-10, Sum P(2) = 1.2e-10
 Identities = 52/141 (36%), Positives = 79/141 (56%)

Query:    372 DTSKQWDYY-ARREKDRDR-ERDRDRERDRDRDRERERTRERERERDHSPTPSVFNSDEE 429
              DT K+ +     ++EK+R   E++R RER+R+R+RERER RERERER+             ++E
Sbjct:   178 DTHKKLEEEKGKKEKERQEIEKER-RERERERERERER-RERERERER-----EREKE 230

Query:    430 RYRYREYAERGYERHRASREKEERHRER---RHREKEETRHKSSRSNSRRRHESEEGDSH 486
              + R RE  ER  +R R         +R RER  R RE+   R+K  RS SR +     E +
Sbjct:   231 KERERE-RERDRDRDRTKERDRDRDRERDRDRDRERSSDRNKD-RSRSREKSRDRERERE 288

Query:    487 RRHKHKKSKRSKEGKEAGSEPAPEQE 512
              R  + ++ +  + +E   E  E+E
Sbjct:   289 RERERERERERERERERERERERE 314

 Score = 171 (25.7 bits), Expect = 2.5e-09, Sum P(2) = 2.5e-09
 Identities = 49/150 (32%), Positives = 78/150 (52%)

Query:    383 REKDRDRERDRDRERDRDRDRERERTRERERERDHSPTPSVFNSDEERYRYREYAERGYE 442
              RE++R+RER+R+RER+R+R+RERER RERERER+              +E+ Y R+  + E
Sbjct:   285 RERERERERERERERERERERERERERERERERKDKKRDREEDEEDAYERRKLERKLRE 344

Query:    443 RHRASREK------EERHRERRHR---EKEETRHKSSRSNSRRRHES-EEGDSHRRH-KH 491
              +  A +E+          ER + R +   E+EE R +      ++R  E  E+ D  R  K+
Sbjct:   345 KEAAYQERLKNWEIRERKKTREYEKEAEREEERRREMAKEAKRLKEFLEDYDDDRDDPKY 404
```

```
Query:   492 ------KKSKRSKEGKEAGSEPAPEQESTE 515
               +K  R +E +    E   ++E  E
Sbjct:   405 YRGSALQKRLRDREKEMEADERDRKREKEE 434

 Score = 162 (24.3 bits), Expect = 2.4e-08, Sum P(2) = 2.4e-08
 Identities = 45/141 (31%), Positives = 74/141 (52%)

Query:   372 DTSKQWDYYARREKDRDRERDRDRERDRDRDRERERTRERERERDHSPTPSVFNSDEERY 431
             + SK  D + + E+++ ++     +E +++R RERER RERERER            + ER
Sbjct:   172 EISKFRDTHKKLEEEKGKKEKERQEIEKER-RERERERERERERERER--ERERE 228

Query:   432 RYREYAERGYERHRASREKEERHRER-RHREKEETRHKSSRSNSRRRHESEEGDSHRRHK 490
             + +E  ER  ER R     +ER R+R R R+++  R +SS  N  R     E+      R +
Sbjct:   229 KEKE-RERERDRDRDRTKERDRDRDRERDRDRDRERSSDRNKDRSRSREKSRDRERER 287

Query:   491 HKKSKRSKEGKEAGSEPAPEQE 512
              ++ +R +E +E   E   E+E
Sbjct:   288 ERERERE-RERERERERE 308

 Score = 137 (20.6 bits), Expect = 1.2e-05, Sum P(2) = 1.2e-05
 Identities = 48/152 (31%), Positives = 68/152 (44%)

Query:   364 APSWPSLVDTSKQWDYYARREKDRDR-ERDRDRERDRDRDRERERTRERERERDHSPTPS 422
             AP  P  + T +  +      E RD    R+  + RD  +  E E+ + +E+ER
Sbjct:   143 APLIPYPLITKEDINAIEMEEDKRDLISREISKFRDTHKKLEEEKGK-KEKERQEIEKER 201

Query:   423 VFNSDEERYRYREYAERGYERHRA-SREKE-ERHRER-RHREKEETRHKS-SRSNSRRRH 478
                  + ER R RE   ER  ER R    REKE ER RER R R+++ T+ +     R   R R
Sbjct:   202 R-ERERERERERERERERERERERERERKEKERERERERDRDRDRTKERDRDRDRERDRD 260

Query:   479 ESEEGDSHRRHKHKKSKRSKEGKEAGSEPAPEQE 512
               E  S R      +S+       +E   E   E+E
Sbjct:   261 RDRERSSDRNKDRSRSREKSRDRERERERERERE 294

 Score = 126 (18.9 bits), Expect = 1.8e-04, Sum P(2) = 1.8e-04
 Identities = 41/149 (27%), Positives = 66/149 (44%)

Query:   375 KQWDYYARREKDRDRERDRDRERDRDRDRERERTRERERERDHSPT---PSVFNSD--EE 429
             K W+     R+K R+ E++ +RE +R R+  +E  R +E   D+     P  +     ++
Sbjct:   354 KNWEI-RERKKTREYEKEAEREEERRREMAKEAKRLKEFLEDYDDDRDDPKYYRGSALQK 412

Query:   430 RYRYREYAERGYERHRASREKEERHRERR------HREKEETRHKSSRSNSRRRHES--E 481
             R R RE        ER R  REKEE       R+       H + +       + +     RRR      +
Sbjct:   413 RLRDREKEMEADERDR-KREKEELEEIRQRLLAEGHPDPDAELQRMEQEAERRRQPQIKQ 471

Query:   482 EGDSHRRHKHKKSKRSKEGKEAGSEPAPEQE 512
             E +S     + K+ K    K   +     E  PEQ+
Sbjct:   472 EPESEEEEEKQEKEEKKREEPMEEEEEPEQK 502

 Score = 124 (18.6 bits), Expect = 3.0e-04, Sum P(2) = 3.0e-04
 Identities = 41/141 (29%), Positives = 65/141 (46%)

Query:   380 YARREKDRD-RERDRDRERDRDRDRERERTRERERERDHSPTPSVFNSDEERYRYREYAE 438
             Y  R K+ + RER + RE +++ +RE ER RE  +E            + + D++R   + Y
Sbjct:   349 YQERLKNWEIRERKKTREYEKEAEREEERRREMAKEAKRLKE-FLEDYDDDRDDPKYYRG 407

Query:   439 RGYERHRASREKEERHRER-RHREKEETRHKSSRSNSRRRHESEEGDSHRRHKHKKSKRS 497
                ++      REKE      ER R REKEE          R +    H   + +  R  +   + +R
Sbjct:   408 SALQKRLRDREKEMEADERDRKREKEELEEIRQRLLAEG-HPDPDAELQRMEQEAERRRQ 466

Query:   498 KEGKEAGSEPAPEQESTEATPAE 520
              + K+    EP  E+E  E     E
Sbjct:   467 PQIKQ---EPESEEEEEKQEKE 486

 Score = 121 (18.2 bits), Expect = 6.2e-04, Sum P(2) = 6.2e-04
 Identities = 43/149 (28%), Positives = 67/149 (44%)

Query:   364 APSWPSLVDTSKQWDYYARREKDRDR-ERDRDRERDRDRDRERERTRERERERDHSPTPS 422
             AP  P  + T +  +      E RD    R+  + RD  +  E E+ + +E+E+ER
Sbjct:   143 APLIPYPLITKEDINAIEMEEDKRDLISREISKFRDTHKKLEEEKGK-KEKERQEIEKE- 200

Query:   423 VFNSDEERYRYREYAERGYERHRASREKEERHRERRHREKEETRHKSSRSNSRRRHESEE 482
                  + ER R RE    R  ER R   RE+E     + R RE+E  R +  R+  R R    E
Sbjct:   201 --RRERERERERERERERER-ERERERERERERERDRDRD-RTKERDRDRDRE 256

Query:   483 GDSHRRHKHKKSKRSKEGKEAGSEPAPEQE 512
               D  R   + +  S R+K+ +      E + ++E
Sbjct:   257 RDRDR-DRERSSDRNKD-RSRSREKSRDRE 284

 Score = 105 (15.8 bits), Expect = 3.1e-02, Sum P(2) = 3.1e-02
```

```
 Identities = 25/73 (34%), Positives = 33/73 (45%)

Query:    428 EERYRYREYAERGYERHRASREKE-ERHRERRHREKEETRHKSSRSNSRRRHESEEGDSH 486
              EE     +E    +  E+ R  RE+E ER RERR RE+E  R +         R  E E
Sbjct:    184 EEEKGKKEKERQEIEKERRERERERERERERRERERERERERERERKEKERERERERDRDR 243

Query:    487 RRHKHKKSKRSKE 499
               R K +    R +E
Sbjct:    244 DRTKERDRDRDRE 256

 Score = 105 (15.8 bits), Expect = 3.1e-02, Sum P(2) = 3.1e-02
 Identities = 31/87 (35%), Positives = 45/87 (51%)

Query:    382 RREKDRDRERDRDRERDRDRDRER-ERTRERERERDHSPTPSVFNSDEERYRYREYAERG 440
              +R +DR++E + D ERDR R++E  E  R+R      H P P       D E  R  + AER
Sbjct:    412 KRLRDREKEMEAD-ERDRKREKEELEEIRQRLLAEGH-PDP-----DAELQRMEQEAERR 464

Query:    441 YERHRASREKEERHRERRHREKEETRHK 468
               + +  +E E     E     +EKEE R +
Sbjct:    465 -RQPQIKQEPESEEEEEEKQEKEEKREE 491

 Score = 46 (6.9 bits), Expect = 1.5e-16, Sum P(2) = 1.5e-16
 Identities = 13/49 (26%), Positives = 21/49 (42%)

Query:     54 AENGVPKPKVTETEDDSDSDSDDDEDDVHVTIGDIKTGAPQYGSYGTAP 102
              A NG  +P+     +D+    D +     D   +  G I+     +Y S    AP
Sbjct:     70 ASNGNARPETVTNDDEEALDEETKRRDQMIK-GAIEVLIREYSSELNAP 117

 Score = 46 (6.9 bits), Expect = 1.8e-04, Sum P(2) = 1.8e-04
 Identities = 14/53 (26%), Positives = 21/53 (39%)

Query:     30 ENEVERPEEENASANPPSGIEDETAENGVPKPKVTETEDDSDSDSDDDEDDVH 82
              + E ER  E               E E  E     + +  E E D   D ++DE+D +
Sbjct:    282 DRERERERERERERERERERER-ERERERERERERKDKKRDREEDEEDAY 333

 Score = 44 (6.6 bits), Expect = 2.0e-13, Sum P(2) = 2.0e-13
 Identities = 13/60 (21%), Positives = 21/60 (35%)

Query:     20 DEEEEWLYGDENEVERPEEENASANPPSGIEDETAENGVPKPKVTETEDDSDSDSDDDED 79
              ++E + +  +  E ER  E                  +   E    K + E E D D D   + D
Sbjct:    191 EKERQEIEKERRERERERERERERRERERERERERERKEKERERERERDRDRDRTKERD 250
```

Pedant information for DKFZphutel_17k7, frame 3

Report for DKFZphutel_17k7.3

```
[LENGTH]        520
[MW]            58375.30
[pI]            5.41
[HOMOL]         PIR:S62454 hypothetical protein SPAC22G7.10 - fission yeast
(Schizosaccharomyces pombe) 3e-18
[FUNCAT]        04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
cerevisiae, YJR093c] 2e-13
[FUNCAT]        30.10 nuclear organization    [S. cerevisiae, YJR093c] 2e-13
[PROSITE]       MYRISTYL        9
[PROSITE]       AMIDATION       1
[PROSITE]       CK2_PHOSPHO_SITE       18
[PROSITE]       TYR_PHOSPHO_SITE       2
[PROSITE]       PKC_PHOSPHO_SITE       12
[PROSITE]       ASN_GLYCOSYLATION      2
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY    35.00 %

SEQ     MSAGEVERLVSELSGGTGGDEEEEWLYGDENEVERPEEENASANPPSGIEDETAENGVPK
SEG     ............................xxxxxxxxxx.....................
PRD     ccccchhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc

SEQ     PKVTETEDDSDSDSDDDEDDVHVTIGDIKTGAPQYGSYGTAPVNLNIKTGGRVYGTTGTK
SEG     ...xxxxxxxxxxxxxxxxx........................................
PRD     cceeeecccccccccccccccceeeeeccccccccccccccccccceeeeeecccceeecccccc

SEQ     VKGVDLDAPGSINGVPLLEVDLDSFEDKPWRKPGADLSDYFNYGFNEDTWKAYCEKQKRI
SEG     ............................................................
PRD     ceeeccccccccccceeeeccccccccccccccccccccccccccccchhhhhhhhhhh

SEQ     RMGLEVIPVTSTTNKITVQQGRTGNSEKETALPSTKAEFTSPPSLFKTGLPPSRRLPGAI
SEG     ............................................................
```

```
PRD     hhhheeeeecccccceeeeeeeecccccccccccccceeeeeccccceeeeecccccccccccc

SEQ     DVIGQTITISRVEGRRRANENSNIQVLSERSATEVDNNFSKPPPFFPPGAPPTHLPPPPF
SEG     ..........................................xxxxxxxxxxxxxxxxxx
PRD     ccccceeeeeecccccccccccceeecccccccccccccccccccccccccccccccccc

SEQ     LPPPPTVSTAPPLIPPPGFPPPPGAPPPSLIPTIESGHSSGYDSRSARAFPYGNVAFPHL
SEG     xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.............................
PRD     cccccccccccccccccccccccccccccccccccccccccccccccccccccceeeccc

SEQ     PGSAPSWPSLVDTSKQWDYYARREKDRDRERDRDRERDRDRERERTRERERERDHSPT
SEG     ......................xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx....
PRD     ccccccccceeeccccchhhhhhhhhccccccccccccccchhhhhhhhhhhcccccc

SEQ     PSVFNSDEERYRYREYAERGYERHRASREKEERHRERRHREKEETRHKSSRSNSRRRHES
SEG     .......xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD     ccccccchhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccc

SEQ     EEGDSHRRHKHKKSKRSKEGKEAGSEPAPEQESTEATPAE
SEG     xx..xxxxxxxxxxxxxx......................
PRD     cccccccccccccccccccccccccccccccccccccccc
```

Prosite for DKFZphutel_17k7.3

PS00001	40->44	ASN_GLYCOSYLATION	PDOC00001
PS00001	278->282	ASN_GLYCOSYLATION	PDOC00001
PS00005	169->172	PKC_PHOSPHO_SITE	PDOC00005
PS00005	193->196	PKC_PHOSPHO_SITE	PDOC00005
PS00005	206->209	PKC_PHOSPHO_SITE	PDOC00005
PS00005	214->217	PKC_PHOSPHO_SITE	PDOC00005
PS00005	233->236	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	373->376	PKC_PHOSPHO_SITE	PDOC00005
PS00005	469->472	PKC_PHOSPHO_SITE	PDOC00005
PS00005	474->477	PKC_PHOSPHO_SITE	PDOC00005
PS00005	485->488	PKC_PHOSPHO_SITE	PDOC00005
PS00005	494->497	PKC_PHOSPHO_SITE	PDOC00005
PS00006	2->6	CK2_PHOSPHO_SITE	PDOC00006
PS00006	17->21	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00006	66->70	CK2_PHOSPHO_SITE	PDOC00006
PS00006	70->74	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	74->78	CK2_PHOSPHO_SITE	PDOC00006
PS00006	84->88	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	215->219	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00006	426->430	CK2_PHOSPHO_SITE	PDOC00006
PS00007	434->442	TYR_PHOSPHO_SITE	PDOC00007
PS00007	152->161	TYR_PHOSPHO_SITE	PDOC00007
PS00008	15->21	MYRISTYL	PDOC00008
PS00008	96->102	MYRISTYL	PDOC00008
PS00008	115->121	MYRISTYL	PDOC00008
PS00008	130->136	MYRISTYL	PDOC00008
PS00008	154->160	MYRISTYL	PDOC00008
PS00008	229->235	MYRISTYL	PDOC00008
PS00008	244->250	MYRISTYL	PDOC00008
PS00008	289->295	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00009	253->257	AMIDATION	PDOC00009

(No Pfam data available for DKFZphutel_17k7.3)

DKFZphutel_18c12
----

group: uterus derived

DKFZphutel_18c12 encodes a novel 378 amino acid protein nearly identical to human WUGSC:H_DJ0872F07.1 protein.

The novel protein has an additional N-terminal domain, which is not present in WUGSC:H_DJ0872F07.1.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

nearly identical to human WUGSC:H_DJ0872F07.1 protein

on genomic level encoded by AC004537, 10 exons the predicted
protein sequence AC004537_1 is only partialy o.k. first exon wasn't
predicted there are additional exons predicted
(BLASTX/EST-BLAST shows that the cDNA is only party spliced)
intron ~1216-3540//~3577-5059

Sequenced by AGOWA

Locus: map="7q31"

Insert length: 6005 bp
Poly A stretch at pos. 5980, polyadenylation signal at pos. 5968

```
   1 AGCGGGTGCT GCTAGCGGAG GCGCCATATT GGAGGGGACA AAACTCCGGC
  51 GACAGCGAGT GACACAAATA AACCCCTGGA CCCCCTTGTT CCCTCAGCTC
 101 TAAGGGCCGC GATGTTGTAC CTAGAAGACT ATCTGGAAAT GATTGAGCAG
 151 CTTCCTATGG ATCTGCGGGA CCGCTTCACG GAAATGCGCG AGATGGACCT
 201 GCAGGTGCAG AATGCAATGG ATCAACTAGA ACAAAGAGTC AGTGAATTCT
 251 TTATGAATGC AAAGAAAAAT AAACCTGAGT GGAGGGAAGA GCAAATGGCA
 301 TCCATCAAAA AAGACTACTA TAAAGCTTTG GAAGATGCAG ATGAGAAGGT
 351 TCAGTTGGCA AACCAGATAT ATGACTTGGT AGATCGACAC TTGAGAAAGC
 401 TGGATCAGGA ACTGGCTAAG TTTAAAATGG AGCTGGAAGC TGATAATGCT
 451 GGAATTACAG AAATATTAGA GAGGCGATCT TTGGAATTAG ACACTCCTTC
 501 ACAGCCAGTG AACAATCACC ATGCTCATTC ACATACTCCA GTGGAAAAAA
 551 GGAAATATAA TCCAACTTCT CACCATACGA CAACAGATCA TATTCCTGAA
 601 AAGAAATTTA AATCTGAAGC TCTTCTATCC ACCCTTACGT CAGATGCCTC
 651 TAAGGAAAAT ACACTAGGTT GTCGAAATAA TAATTCCACA GCCTCTTCTA
 701 ACAATGCCTA CAATGTGAAT TCCTCCCAAC CTCTGGGATC CTATAACATT
 751 GGCTCGTTAT CTTCAGGAAC TGGTGCAGGG GCAATTACCA TGGCAGCTGC
 801 TCAAGCAGTT CAGGCTACAG CTCAGATGAA GGAGGGACGA AGAACATCAA
 851 GTTTAAAAGC CAGTTATGAA GCATTTAAGA ATAATGACTT TCAGTTGGGA
 901 AAAGAATTTT CAATGGCCAG GGAAACAGTT GGCTATTCAT CATCTTCGGC
 951 ACTTATGACA ACATTAACAC AGAATGCCAG TTCATCAGCA GCCGACTCAC
1001 GGAGTGGTCG AAAGAGCAAA AACAACAACA AGTCTTCAAG CCAGCAGTCA
1051 TCATCTTCCT CCTCCTCTTC TTCCTTATCA TCGTGTTCTT CATCATCAAC
1101 TGTTGTACAA GAAATCTCTC AACAAACAAC TGTAGTGCCA GAATCTGATT
1151 CAAATAGTCA GGTTGATTGG ACTTACGACC CAAATGAACC TCGATACTGC
1201 ATTTGTAATC AGGTAAAAGT CTGTTATATC TATAAAAGTA TAATCTGAAT
1251 AAACTAGAAG GAAGAGAACT ATTTCATTTT TAAGCACTTT TTTAAACTCA
1301 CTTAAAATAC CTTTGCTTTA TTTGTATACT TTTCTCCCCC TTCTTACAAA
1351 AGTGACATTT GCTGTAAATA CTGAGTATAA AGAAAAATGT TACCCATAAT
1401 CCTAGCCCTC AGATACAACC TGTAACTAAA CATTTTTGGT ATACCACTAC
1451 CATATACCTC ATGTGCACAT TGGCTGCCTT AATAAAATAC AACAGACTGG
1501 GTAGCTTAAA CAACAGAAAA TAATTTTCTC ACAGGTATGA AGGCTGGGAA
1551 GTCCAAGATC AAGGTGTCCA CTGACTCAGT TCTGGAGGAG GGCTCCCTTC
1601 CTAGATGGAG ACTGCTGCCT TCTCACCGGG TCCTCACATG ATAGAGGGAG
1651 AAAGAGTGTG CTCTGGTGTC TTTTCTTATA AGGGCACCAG CCTTGTCAGA
1701 GTAGGACCCC ACTCTATGAC CTCATTTAAC CTTTACCACC TCCTCACAGG
1751 CCCTGTTTCC AATTATAGTC ACGTTGGGGG TTAGGGCTTC AACATATGAT
1801 TTTGAGACAT AAGCTTGCAT TTCATAACAC GTGTCTATGC AGATTTGCAC
1851 ATGCATGTGT GTATAAGTTT GTCAGTAGGA ACCACAGTGT ATACTTTCTT
1901 GTTACTGGCT TTTTTCTCTA AATCAGGTAT ACCGAACATG ATTTTTCTTT
1951 AAGATCATAT TTTTAATTTT CACATAGTTA TCTCTTATGC CATCCAGTGT
2001 AGTTTTCTTA ACCAATACCT AGCTATAGAT TATATTAGTG GTTTTAATTT
2051 GTTTGAAATT AGGGATAATA TTACGATAGG CATTTTTTAA ATGTAATCCA
2101 TTTTATACAT CTAATTTCTT GGATAATCTT TTAGAAATAA AATTAGGCTG
2151 TAAATATTTG ACAGACACCA AAATATATTT TCTAGAAATT TATTACCAAA
2201 AATTAATAAA CATACCGGTT TACTAAACCC TGTCCAACAC TGGATATTAT
2251 TTTCTTTTAA AAACTAAGTA CCAATTTGGT AGTTTTATAT TATGATTGTT
2301 TTAAATACAC TAGTATTATT GAAGTTGGAC ATTTTTTGAC CATTTTTGTT
2351 TTTTACATTA TGAATCGACT CCTAATGGTG TCGGCTGATT TTTCTATTGT
```

```
2401 TTTTGTTATG TACTCTAAAT ATTTGCTTGA TTTAGTTTTT TAAAAATAAT
2451 TCTAAAATTT TAATTTTATG TAGTTATGAC TGTTAATTTT TTTTTATGAA
2501 GCAAGCCATG GATTATATAC TTAGAAGGGC TTTCTCTTTG GCTCTTCTTT
2551 CTACAAAAAA TTGTCTTGTA TAATATTTTC TCCTAGTTTT TATATGGTTT
2601 TGTCTAGTTC TTTGCATGCT TCAGTTTCTT CACATTTAAG ACTTAGTCTA
2651 TCAGCAGATT ATTGTGTCTA ACAGTATGAG TTGCCAGTCT GATTTTTAAA
2701 AATTTTAACA ATTTGTTAGC TGTTCCACTA TCACCCGATA AACATTTTTC
2751 AGTACAAATG ATAGAAAAGC ATATCCTGTA TCCTGACAAC AAAAGTAGAT
2801 TACTTGCAAA AGAACAAAAT CAGACTGAAC CTAGAGTTTT CCTCTGTAAC
2851 ACTAAAAAAC TAGAAGGTGA TGGAATATGT CTGTAGAGCT TTCAGGGAAA
2901 AATTAAGAGC CCCCAAAAAC TTGATATTCA GAGAAGTTAT TTCTCTGCAT
2951 AGGACCATGT AAATATATTT TCACTCATGC AGAGAATCAG AAGATATGCC
3001 ATCTAGTTAA TCCTGTCTGA AAAATTATTC AATCCACTGA GAACTTCAGT
3051 GAACTCAAGA ATTAGCAAGT TATGCCCTAA AGTGCTGGTG ATGAAGAGCA
3101 AAAGAAAAAT GAGAAAGGAC ATAAAATAGA TAAGTTTAGA AGTTTCAAGG
3151 AAGGAGACTA TTAATTGCAA AAATATATAT GACCTAATGT GACCCAAGAA
3201 GTAAAAACTT TCAGTAAGTA AATAATCAAG AAAGGAACTT AAAATTTTTA
3251 CAATAAGAAC TACCCAGAAA GATGACTCCT TCATCCGGGT GATTTATATG
3301 TCAAGTTCTT CCAGACTTCT GAAGGGCAGA TAATTCCTGT GCATTTCTTC
3351 CCACCCTTGC CCCACCCTGC CCAAAAGAGT ATTTCAGGAA AAAATTATTA
3401 TACCTTGATT CTCAATGTAA TTGTATATTC AGTGTATTTC CCTTTATTTT
3451 CCAGCAGTAT CATACATAAA CAGTTAATTG GTATCTAGGT GTTTGTTACA
3501 TAGTCATAAT AAAGACATTT AATTTTTTTT AACTAGGTAT CTTATGGTGA
3551 GATGGTGGGA TGTGATAACC AAGATGTAAG TATTACATTT TTCTATTTAG
3601 GAATGAAAAA AATCACAGGT TGTTATTACT TGAATATTTG TCTTATTTGC
3651 TGTATGGTTT GGTCTAAGAA AACAGGTTTG CAGGTATATT AGTTATGTTA
3701 TGCTAATGCT AGAATATTCC TCTTCAAAAT AGGGTAGTGT CCCTTAATGT
3751 GTTCCCTATT TTAATTTTTA AAGCTAATTT TATGGTTTTA TGTGCAGATT
3801 GTCTCAGAAG TGTTATGTTG TATGAAAATT ATAAATACCC TCCTTTCCCT
3851 TTACTAAAAA ATACTGTGTT TACTAGAATC CAGTTCATTT ATCACATTGA
3901 AGAAATGGAA TTTTAAAACA ATTCATTCTT TCAGGCTGCA CCGTGCTAAA
3951 GTGAAGGGTG GGATAATTGA GGATCTAATG TGAGATTATC TTCCTCTCAT
4001 GAGTATAATA TTTTTTCCTG TACTCTGCAG GTGTCAGCTG ATAAGAGCCA
4051 CCCCTGATCT AAAAAGTAAA GGAAATTTGA AAGGAAGGAA TTCTTGGTTT
4101 TTAGGAGACT TAATTTTAGT TAGAGATACG TTTTTTATTC AATACTGAGA
4151 ATATTGTTGT CTAGTAATTT TGACTCCCTC CTTATTTAGT AGTGACAGGA
4201 TCCTAAGATT AACAAGAGTT TTAAATTTGT AAAACAATCT GAAGATTGAG
4251 GGAGCTGGCT AGGTGCATTA AAATGTGTAC TTTTCCTAGA CCTGATAGGG
4301 TTACAGCAAC ATGCTCACGT AGATTGGGAC AGAGCCTCCT TCTGTTTCCC
4351 TGTCTAGAAT CCCTTGTAGG CTGTTTGTGG TTGTTGCAAA AACAATATTG
4401 CCCAACCATT TCAAGAACAT CACTGTAAAC TCTTCTGGGG CAGTTAGTGA
4451 AAATGATGAA TGAGATTTCT ATGAGTACCA GCATCATGCT TCTCTGATTC
4501 TTCTTATTCC CAGTTGTGCT CTTCTGAGTG CTAAGACTTT CATGAAAGAG
4551 TTTTCTGCTT AATATGTTTC AAAGAGGAAT AATTTTTCTC TACATTTCAA
4601 GGAATAGAAA CACCCACGTA GGAAATGCAG GGCATAAGAC ATAAATTAAT
4651 GTCTTTAATT ACAATCAGCT TATTCTACTT TATGAGACAG CAAATAAGGC
4701 TGACTATTAA ATAAAATCTT AAGTTATATT TACCTTCTAC ATAGAAGATT
4751 CATCCCACTT CTTTTTGCCC TTGAAAGCTG AAAACTAGTG AATTTTCATT
4801 CATTAGGATG AGGGGACTAG ATTACATGGA CCTCAGGATT CTTGAAGATG
4851 CATAATTTTT CTGTGCCTTC ATTTCCTCAT TCCTGAAGCT TATCATTTAG
4901 TCTAAATGAT GTCTAAATAA TCTAGATCTA AAAATTCTGA TGTCACACAT
4951 CTAATTATTG TTAAATTAAA TGGATTATTC AGTCTCCTGA GCATATTTTA
5001 ATATACTCTC TTGTCTTCAG AAGTACTGAA AACTTGTTTT TTGCAATTTT
5051 GCTTTCTAGT GCCCTATAGA ATGGTTCCAT TATGGCTGCG TTGGATTGAC
5101 AGAGGCACCA AAAGGCAAAT GGTACTGTCC ACAGTGCACT GCTGCAATGA
5151 AGAGAAGAGG CAGCAGACAC AAATAAAGGT GGTCCTTTTG TTTGATGAAG
5201 AAATAAACTT CAGCTGAAGA TTTTATATAG GACTTTAAAA AGAAGAGAAG
5251 AGAAAGAAGA AACAATGCAT TTCCAGGCAA CCACTTAAAG GATTTACATA
5301 GACAATCCTA TAAGATCTTG AACTTGAATT TTATGGGTTG TATTTTAATA
5351 ATGTAAGTAA ATTATTTATG CACTCCTGGT GTGCTATGAA TATTATTCCA
5401 GTTAGCCTTG GATTATTTCA GTGGCCAACA TATGCAGACA TTTGTACTCC
5451 TCAACCATTT TCTCAAAGTA ATGGGCATTC TATGATTTAG ACTTCAAGGA
5501 ATTCCAATGA TGAAGATTTT AAGGAAAGTA TTTTATATTC AACAGGTATA
5551 TTCTGCTGCA TGTACTGTAC TCCAGAGCTG TTATGTAACA CTGTATATAA
5601 ATGGTTGCAA AAAAAAAAAA AAGTCAGTGC TTCTAAAAAG AATTTAAGAT
5651 AATGGTTTTT AAAATGCCTT TATAATAAGC TTTGTTTCTT TGTGAAACTA
5701 ATTCAGCAGG CTGAAGGAAA TGGTTCATGT GATAATGTGG GCTGGTATCC
5751 TCTAGAGTAC CTGGGTACAT AAACAGAAAC TCCTGTAGGT AAAAAGTAAT
5801 TTGTGCCATT AGTCTTTCTA TGTTTCTGCA TCCAGATAGA GTGCAGTTCA
5851 TGAGGGAGGG GGCGGGGGAC TGAAGGGGAA AGGGCGTTAA AGTGATACAT
5901 TTTTATACCA AATGTGTTTA TTTTTTTGTG CAAGTAATCC TTAAAATTGC
5951 AATTGTATTA GGTGTTAAAA TAAAGTTTTT AAAAAATTAA AAAAAAAAAA
6001 AAAAA
```

BLAST Results

```
Entry HSG20547 from database EMBL:
HSG20547| human STS A005W09.
Length = 154
```

```
Minus Strand HSPs:
Score = 770 (115.5 bits), Expect = 2.9e-26, P = 2.9e-26
Identities = 154/154 (100%)
```

Medline entries

```
98101645:
The candidate tumour suppressor p33ING1 cooperates with p53 in cell
growth control.
```

Peptide information for frame 1

```
ORF from 112 bp to 1245 bp; peptide length: 378
Category: similarity to known protein

    1 MLYLEDYLEM IEQLPMDLRD RFTEMREMDL QVQNAMDQLE QRVSEFFMNA
   51 KKNKPEWREE QMASIKKDYY KALEDADEKV QLANQIYDLV DRHLRKLDQE
  101 LAKFKMELEA DNAGITEILE RRSLELDTPS QPVNNHHAHS HTPVEKRKYN
  151 PTSHHTTTDH IPEKKFKSEA LLSTLTSDAS KENTLGCRNN NSTASSNNAY
  201 NVNSSQPLGS YNIGSLSSGT GAGAITMAAA QAVQATAQMK EGRRTSSLKA
  251 SYEAFKNNDF QLGKEFSMAR ETVGYSSSSA LMTTLTQNAS SSAADSRSGR
  301 KSKNNNKSSS QQSSSSSSSS SLSSCSSSST VVQEISQQTT VVPESDSNSQ
  351 VDWTYDPNEP RYCICNQVKV CYIYKSII
```

BLASTP hits

```
Entry AF044076_1 from database TREMBL:
"ING1"; product: "candidate tumor suppressor p33ING1";      Homo
sapiens candidate tumor suppressor p33ING1 (ING1) mRNA, complete
cds. Homo sapiens (human)
Length = 279
Score = 162 (57.0 bits), Expect = 1.1e-09, P = 1.1e-09
Identities = 48/183 (26%), Positives = 92/183 (50%)

Entry AC004537_1 from database TREMBL:
gene: "WUGSC:H_DJ0872F07.1"; Homo sapiens PAC clone DJ0872F07 from
7q31, complete sequence.
Score = 1814, P = 3.7e-187, identities = 358/358, positives = 358/358

Entry CEY51H1A_1 from database TREMBL:
gene: "Y51H1A.4"; Caenorhabditis elegans cosmid Y51H1A
Score = 213, P = 3.7e-15, identities = 37/123, positives = 82/123
```

Alert BLASTP hits for DKFZphutel_18c12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphutel_18c12, frame 1

Report for DKFZphutel_18c12.1

```
[LENGTH]        378
[MW]            42275.72
[pI]            5.72
[HOMOL]         TREMBL:AC004537_1 gene: "WUGSC:H_DJ0872F07.1"; Homo sapiens PAC clone DJ0872F07
from 7q31, complete sequence. 1e-157
[FUNCAT]        99 unclassified proteins       [S. cerevisiae, YHR090c] 8e-05
[FUNCAT]        04.05.01.04 transcriptional control   [S. cerevisiae, YNL097c] 2e-04
[PROSITE]       MYRISTYL        3
[PROSITE]       AMIDATION       2
[PROSITE]       CAMP_PHOSPHO_SITE       1
[PROSITE]       CK2_PHOSPHO_SITE        4
[PROSITE]       PROKAR_LIPOPROTEIN      1
[PROSITE]       GLYCOSAMINOGLYCAN       1
[PROSITE]       PKC_PHOSPHO_SITE        3
[PROSITE]       ASN_GLYCOSYLATION       5
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY    20.63 %
```

```
[KW]          COILED_COIL         7.94 %

SEQ     MLYLEDYLEMIEQLPMDLRDRFTEMREMDLQVQNAMDQLEQRVSEFFMNAKKNKPEWREE
SEG     ............................................................
PRD     ccchhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhh
COILS   ............................................................

SEQ     QMASIKKDYYKALEDADEKVQLANQIYDLVDRHLRKLDQELAKFKMELEADNAGITEILE
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhh
COILS   ............CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC..................
        -
SEQ     RRSLELDTPSQPVNNHHAHSHTPVEKRKYNPTSHHTTTDHIPEKKFKSEALLSTLTSDAS
SEG     ............................................................
PRD     hhcccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhcccc
COILS   ............................................................

SEQ     KENTLGCRNNNSTASSNNAYNVNSSQPLGSYNIGSLSSGTGAGAITMAAAQAVQATAQMK
SEG     ........xxxxxxxxxxxxxxxxx..............xxxxxxxxxxxxxxxxxxx..
PRD     cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhh
COILS   ............................................................

SEQ     EGRRTSSLKASYEAFKNNDFQLGKEFSMARETVGYSSSSALMTTLTQNASSSAADSRSGR
SEG     ................................................xxxxxxxxxxxx
PRD     hcccccccchhhhhcccccccccccccccccccccccceeeeecccccccccccccccccc
COILS   ............................................................

SEQ     KSKNNNKSSSQQSSSSSSSSSLSSCSSSSTVVQEISQQTTVVPESDSNSQVDWTYDPNEP
SEG     xxxxxxxxxxxxxxxxxxxxxxxxxxxx................................
PRD     ccccccccccccccccccccceeecccccccccccccccccccccccccceeeecccccc
COILS   ............................................................

SEQ     RYCICNQVKVCYIYKSII
SEG     ..................
PRD     eeeeceeeeeeeeeeccc
COILS   ..................
```

Prosite for DKFZphutel_18c12.1

PS00001	190->194	ASN_GLYCOSYLATION	PDOC00001
PS00001	191->195	ASN_GLYCOSYLATION	PDOC00001
PS00001	203->207	ASN_GLYCOSYLATION	PDOC00001
PS00001	288->292	ASN_GLYCOSYLATION	PDOC00001
PS00001	306->310	ASN_GLYCOSYLATION	PDOC00001
PS00002	218->222	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	243->247	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	64->67	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	298->301	PKC_PHOSPHO_SITE	PDOC00005
PS00006	142->146	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	349->353	CK2_PHOSPHO_SITE	PDOC00006
PS00008	186->192	MYRISTYL	PDOC00008
PS00008	214->220	MYRISTYL	PDOC00008
PS00008	219->225	MYRISTYL	PDOC00008
PS00009	241->245	AMIDATION	PDOC00009
PS00009	298->302	AMIDATION	PDOC00009
PS00013	315->326	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphutel_18c12.1)

DKFZphutel_18i19
----

group: transcription factors

DKFZphutel_18i19 encodes a novel 759 amino acid protein with similarity to the SREBP-2 mutant sterol regulatory element binding protein-2 of Cricetulus griseus.

The SREBP-2 protein is embedded in the membranes of the nucleus and endoplasmic reticulum. In cholesterol-depleted cells the proteins are cleaved to release soluble NH2-terminal fragments that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis. The new protein is a putative transcription factor capable of protein-protein interaction via a lim domain and additionally shows similarity to the common sunflower transcription factor SF3.

The new protein can find application in modulating/blocking the expression of genes involved in lipid metabolism.

similarity to transcription factor SF3

complete cDNA, complete cds, EST hits
strong similarity to mutated SREBP-2 of hamster,
similarity is not to SREP-2 part of protein but to the unknown part of the fusion protein

Sequenced by AGOWA

Locus: /map=12

Insert length: 3664 bp
Poly A stretch at pos. 3647, polyadenylation signal at pos. 3636

```
   1 GCGCTAGGTA GAGCGCCGGG ACCTGTGACA GGGCTGGTAG CAGCGCAGAG
  51 GAAAGGCGGC TTTTAGCCAG GTATTTCAGT GTCTGTAGAC AAGATGGAAT
 101 CATCTCCATT TAATAGACGG CAATGGACCT CACTATCATT GAGGGTAACA
 151 GCCAAAGAAC TTTCTCTTGT CAACAAGAAC AAGTCATCGG CTATTGTGGA
 201 AATATTCTCC AAGTACCAGA AAGCAGCTGA AGAAACAAAC ATGGAGAAGA
 251 AGAGAAGTAA CACCGAAAAT CTCTCCCAGC ACTTTAGAAA GGGGACCCTG
 301 ACTGTGTTAA AGAAGAAGTG GGAGAACCCA GGGCTGGGAG CAGAGTCTCA
 351 CACAGACTCT CTACGGAACA GCAGCACTGA GATTAGGCAC AGAGCAGACC
 401 ATCCTCCTGC TGAAGTGACA AGCCACGCTG CTTCTGGAGC CAAAGCTGAC
 451 CAAGAAGAAC AAATCCACCC CAGATCTAGA CTCAGGTCAC CTCCTGAAGC
 501 CCTCGTTCAG GGTCGATATC CCCACATCAA GGACGGTGAG GATCTTAAAG
 551 ACCACTCAAC AGAAAGTAAA AAAATGGAAA ATTGTCTAGG AGAATCCAGG
 601 CATGAAGTAG AAAAATCAGA AATCAGTGAA AACACAGATG CTTCGGGCAA
 651 AATAGAGAAA TATAATGTTC CGCTGAACAG GCTTAAGATG ATGTTTGAGA
 701 AAGGTGAACC AACTCAAACT AAGATTCTCC GGGCCCAAAG CCGAAGTGCA
 751 AGTGGAAGGA AGATCTCTGA AAACAGCTAT TCTCTAGATG ACCTGGAAAT
 801 AGGCCCAGGT CAGTTGTCAT CTTCTACATT TGACTCGGAG AAAAATGAGA
 851 GTAGACGAAA TCTGGAACTT CCACGCCTCT CAGAAACCTC TATAAAGGAT
 901 CGAATGGCCA AGTACCAGGC AGCTGTGTCC AAACAAAGCA GCTCAACCAA
 951 CTATACAAAT GAGCTGAAAG CCAGTGGTGG CGAAATCAAA ATTCATAAAA
1001 TGGAGCAAAA GGAGAATGTG CCCCCAGGTC CTGAGGTCTG CATCACCCAT
1051 CAGGAAGGGG AAAAGATTTC TGCAAATGAG AATAGCCTGG CAGTCCGTTC
1101 CACCCCTGCC GAAGATGACT CCCGTGACTC CCAGGTTAAG AGTGAGGTTC
1151 AACAGCCTGT CCATCCCAAG CCACTAAGTC CAGATTCCAG AGCCTCCAGT
1201 CTTTCTGAAA GTTCTCCTCC CAAAGCAATG AAGAAGTTTC AGGCACCTGC
1251 AAGAGAGACC TGCGTGGAAT GTCAGAAGAC AGTCTATCCA ATGGAGCGTC
1301 TCTTGGCCAA CCAGCAGGTG TTTCACATCA GCTGCTTCCG TTGCTCCTAT
1351 TGCAACAACA AACTCAGTCT AGGAACATAT GCATCTTTAC ATGGAAGAAT
1401 CTATTGTAAG CCTCACTTCA ATCAACTCTT TAAATCTAAG GGCAACTATG
1451 ATGAAGGCTT TGGGCACAGA CCACACAAGG ATCTATGGGC AAGCAAAAAT
1501 GAAAACGAAG AGATTTTGGA GAGACCAGCC CAGCTTGCAA ATGCAAGGGA
1551 GACCCCTCAC AGCCCAGGGG TAGAAGATGC CCCTATTGCT AAGGTGGGTG
1601 TCCTGGCTGC AAGTATGGAA GCCAAGGCCT CCTCTCAGCA GGAGAAGGAA
1651 GACAAGCCAG CTGAAACCAA GAAGCTGAGG ATCGCCTGGC CACCCCCCAC
1701 TGAACTTGGA AGTTCAGGAA GTGCCTTGGA GGAAGGGATC AAAATGTCAA
1751 AGCCCAAATG GCCTCCTGAA GACGAAATCA GCAAGCCCGA AGTTCCTGAG
1801 GATGTCGATC TAGATCTGAA GAAGCTAAGA CGATCTTCTT CACTGAAGGA
1851 AAGAAGCCGC CCATTCACTG TAGCAGCTTC ATTTCAAAGC ACCTCTGTCA
1901 AGAGCCCAAA AACTGTGTCC CCACCTATCA GGAAAGGCTG GAGCATGTCA
1951 GAGCAGAGTG AAGAGTCTGT GGGTGGAAGA GTTGCAGAAA GGAAACAAGT
2001 GGAAAATGCC AAGGCTTCTA AGAAGAATGG GAATGTGGGA AAAACAACCT
2051 GGCAAAACAA AGAATCTAAA GGAGAGACAG GGAAGAGAAG TAAGGAAGGT
2101 CATAGTTTGG AGATGGAGAA TGAGAATCTT GTAGAAAATG GTGCAGACTC
2151 CGATGAAGAT GATAACAGCT TCCTCAAACA ACAATCTCCA CAAGAACCCA
2201 AGTCTCTGAA TTGGTCGAGT TTTGTAGACA ACACCTTTGC TGAAGAATTC
2251 ACTACTCAGA ATCAGAAATC CCAGGATGTG GAACTCTGGG AGGGAGAAGT
```

```
2301 GGTCAAAGAG CTCTCTGTGG AAGAACAGAT AAAGAGAAAT CGGTATTATG
2351 ATGAGGATGA GGATGAAGAG TGACAAATTG CAATGATGCT GGGCCTTAAA
2401 TTCATGTTAG TGTTAGCGAG CCACTGCCCT TTGTCAAAAT GTGATGCACA
2451 TAAGCAGGTA TCCCAGCATG AAATGTAATT TACTTGGAAG TAACTTTGGA
2501 AAAGAATTCC TTCTTAAAAT CAAAAACAAA ACAAAAAAAC ACAAAAAAACA
2551 CATTCTAAAT ACTAGAGATA ACTTTACTTA AATTCTTCAT TTTAGCAGTG
2601 ATGATATGCG TAAGTGCTGT AAGGCTTGTA ACTGGGGAAA TATTCCACCT
2651 GATAATAGCC CAGATTCTAC TGTATTCCCA AAAGGCAATA TTAAGGTAGA
2701 TAGATGATTA GTAGTATATT GTTACACACT ATTTTGGAAT TAGAGAACAT
2751 ACAGAAGGAA TTTAGGGGCT TAAACATTAC GACTGAATGC ACTTTAGTAT
2801 AAAGGGCACA GTTTGTATAT TTTTAAATGA ATACCAATTT AATTTTTTAG
2851 TATTTACCTG TTAAGAGATT ATTTAGTCTT TAAATTTTTT AGGTTAATTT
2901 TCTTGCTGTG ATATATATGA GGAATTTACT ACTTTATGTC CTGCTCTCTA
2951 AACTACATCC TGAACTCGAC GTCCTGAGGT ATAATACAAC AGAGCACTTT
3001 TTGAGGCAAT TGAAAAACCA ACCTACACTC TTCGGTGCTT AGAGAGATCT
3051 GCTGTCTCCC AAATAAGCTT TTGTATCTGC CAGTGAATTT ACTGTACTCC
3101 AAATGATTGC TTTCTTTTCT GGTGATATCT GTGCTTCTCA TAATTACTGA
3151 AAGCTGCAAT ATTTTAGTAA TACCTTCGGG ATCACTGTCC CCCATCTTCC
3201 GTGTTAGAGC AAAGTGAAGA GTTTAAAGGA GGAAGAAGAA AGAACTGTCT
3251 TACACCACTT GAGCTCAGAC CTCTAAACCC TGTATTTCCC TTATGATGTC
3301 CCCTTTTTGA GACACTAATT TTTAAATACT TACTAGCTCT GAAATATATT
3351 GATTTTTATC ACAGTATTCT CAGGGTGAAA TTAAACCAAC TATAGGCCTT
3401 TTTCTTGGGA TGATTTTCTA GTCTTAAGGT TTGGGGACAT TATAAACTTG
3451 AGTACATTTG TTGTACACAG TTGATATTCC AAATTGTATG GATGGGAGGG
3501 AGAGGTGTCT TAAGCTGTAG GCTTTTCTTT GTACTGCATT TATAGAGATT
3551 TAGCTTTAAT ATTTTTTAGA GATGTAAAAC ATTCTGCTTT CTTAGTCTTA
3601 CCTAGTCTGA AACATTTTTA TTCAATAAAG ATTTTAATTA AAATTTGAAA
3651 AAAAAAAAAA AAAA
```

BLAST Results

Entry HS512217 from database EMBL:
human STS SHGC-14654.
Length = 250
Minus Strand HSPs:
Score = 1202 (180.3 bits), Expect = 1.8e-46, P = 1.8e-46
Identities = 242/244 (99%)

Medline entries

95263566:
Three different rearrangements in a single intron truncate sterol regulatory element binding protein-2 and produce sterol-resistant phenotype in three cell lines. Role of introns in protein evolution.

93258417:
Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein.

Peptide information for frame 1

ORF from 94 bp to 2370 bp; peptide length: 759
Category: similarity to known protein

```
  1 MESSPFNRRQ WTSLSLRVTA KELSLVNKNK SSAIVEIFSK YQKAAEETNM
 51 EKKRSNTENL SQHFRKGTLT VLKKKWENPG LGAESHTDSL RNSSTEIRHR
101 ADHPPAEVTS HAASGAKADQ EEQIHPRSRL RSPPEALVQG RYPHIKDGED
151 LKDHSTESKK MENCLGESRH EVEKSEISEN TDASGKIEKY NVPLNRLKMM
201 FEKGEPTQTK ILRAQSRSAS GRKISENSYS LDDLEIGPGQ LSSSTFDSEK
251 NESRRNLELP RLSETSIKDR MAKYQAAVSK QSSSTNYTNE LKASGGEIKI
301 HKMEQKENVP PGPEVCITHQ EGEKISANEN SLAVRSTPAE DDSRDSQVKS
351 EVQQPVHPKP LSPDSRASSL SESSPPKAMK KFQAPARETC VECQKTVYPM
401 ERLLANQQVF HISCFRCSYC NNKLSLGTYA SLHGRIYCKP HFNQLFKSKG
451 NYDEGFGHRP HKDLWASKNE NEEILERPAQ LANARETPHS PGVEDAPIAK
501 VGVLAASMEA KASSQQEKED KPAETKKLRI AWPPPTELGS SGSALEEGIK
551 MSKPKWPPED EISKPEVPED VDLDLKKLRR SSSLKERSRP FTVAASFQST
601 SVKSPKTVSP PIRKGWSMSE QSEESVGGRV AERKQVENAK ASKKNGNVGK
651 TTWQNKESKG ETGKRSKEGH SLEMENENLV ENGADSDEDD NSFLKQQSPQ
701 EPKSLNWSSF VDNTFAEEFT TQNQKSQDVE LWEGEVVKEL SVEEQIKRNR
```

751 YYDEDEDEE

BLASTP hits

Entry CG22818_1 from database TREMBL:
"SREBP-2"; product: "mutant sterol regulatory element binding protein-2"; Cricetulus griseus SRD-2 mutant sterol regulatory element binding protein-2 (SREBP-2) mRNA, complete cds. Cricetulus griseus (Chinese hamster)
Length = 839
Score = 1502 (528.7 bits), Expect = 3.9e-154, P = 3.9e-154
Identities = 290/380 (76%), Positives = 322/380 (84%)

Entry S28507 from database PIR:
transcription factor SF3 - common sunflower
Length = 219
Score = 212 (74.6 bits), Expect = 6.3e-18, Sum P(2) = 6.3e-18
Identities = 36/82 (43%), Positives = 55/82 (67%)

Entry NTLIMDOM_1 from database TREMBL:
"SF3"; product: "LIM-domain SF3 protein"; N.tabacum mRNA for LIM-domain protein Nicotiana tabacum (common tobacco)
Length = 189
Score = 216 (76.0 bits), Expect = 1.0e-16, P = 1.0e-16
Identities = 42/94 (44%), Positives = 57/94 (60%)

Alert BLASTP hits for DKFZphutel_18i19, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphutel_18i19, frame 1

Report for DKFZphutel_18i19.1

[LENGTH] 759
[MW] 85225.57
[pI] 6.41
[HOMOL] TREMBL:CG22818_1 gene: "SREBP-2"; product: "mutant sterol regulatory element binding protein-2"; Cricetulus griseus SRD-2 mutant sterol regulatory element binding protein-2 (SREBP-2) mRNA, complete cds. 1e-151
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR257w] 3e-05
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YGR162w TIF4631 - mRNA cap-binding protein] 1e-04
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YGR162w TIF4631 - mRNA cap-binding protein] 1e-04
[BLOCKS] BL00478B
[PIRKW] zinc finger 9e-16
[PIRKW] DNA binding 9e-16
[SUPFAM] LIM metal-binding repeat homology 9e-16
[PROSITE] MYRISTYL 6
[PROSITE] LIM_DOMAIN_1 1
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 4
[PROSITE] CK2_PHOSPHO_SITE 28
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 15
[PROSITE] ASN_GLYCOSYLATION 6
[PFAM] LIM domain containing proteins
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 5.53 %

```
SEQ    MESSPFNRRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMEKKRSNTENL
SEG    ............................................................
1ctl-  ............................................................

SEQ    SQHFRKGTLTVLKKKWENPGLGAESHTDSLRNSSTEIRHRADHPPAEVTSHAASGAKADQ
SEG    ............................................................
1ctl-  ............................................................

SEQ    EEQIHPRSRLRSPPEALVQGRYPHIKDGEDLKDHSTESKKMENCLGESRHEVEKSEISEN
SEG    ............................................................
1ctl-  ............................................................

SEQ    TDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRSASGRKISENSYSLDDLEIGPGQ
SEG    ............................................................
```

```
1ctl-   ..........................................................

SEQ     LSSSTFDSEKNESRRNLELPRLSETSIKDRMAKYQAAVSKQSSSTNYTNELKASGGEIKI
SEG     ............................................................
1ctl-   ............................................................

SEQ     HKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVKSEVQQPVHPKP
SEG     ...........................................................x
1ctl-   ............................................................

SEQ     LSPDSRASSLSESSPPKAMKKFQAPARETCVECQKTVYPMERLLANQQVFHISCFRCSYC
SEG     xxxxxxxxxxxxxxx.............................................
1ctl-   ..............................ETTTTEEETTTCEEEETTEEEETTTTBTTTT

SEQ     NNKLSLGTYASLHGRIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEEILERPAQ
SEG     ............................................................
1ctl-   TCBCBTTBEEEETTEEEETTTTTTTTTCCTTTTTTTCTTT....................

SEQ     LANARETPHSPGVEDAPIAKVGVLAASMEAKASSQQEKEDKPAETKKLRIAWPPPTELGS
SEG     ............................................................
1ctl-   ............................................................

SEQ     SGSALEEGIKMSKPKWPPEDEISKPEVPEDVDLDLKKLRRSSSLKERSRPFTVAASFQST
SEG     ..............................xxxxxxxxxxxxxxxxx............
1ctl-   ............................................................

SEQ     SVKSPKTVSPPIRKGWSMSEQSEESVGGRVAERKQVENAKASKKNGNVGKTTWQNKESKG
SEG     ............................................................
1ctl-   ............................................................

SEQ     ETGKRSKEGHSLEMENENLVENGADSDEDDNSFLKQQSPQEPKSLNWSSFVDNTFAEEFT
SEG     ............................................................
1ctl-   ............................................................

SEQ     TQNQKSQDVELWEGEVVKELSVEEQIKRNRYYDEDEDEE
SEG     ................................xxxxxxx
1ctl-   .......................................
```

Prosite for DKFZphutel_18i19.1

PS00001	29->33	ASN_GLYCOSYLATION	PDOC00001
PS00001	59->63	ASN_GLYCOSYLATION	PDOC00001
PS00001	92->96	ASN_GLYCOSYLATION	PDOC00001
PS00001	251->255	ASN_GLYCOSYLATION	PDOC00001
PS00001	286->290	ASN_GLYCOSYLATION	PDOC00001
PS00001	706->710	ASN_GLYCOSYLATION	PDOC00001
PS00004	52->56	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	65->69	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	222->226	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	579->583	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	15->18	PKC_PHOSPHO_SITE	PDOC00005
PS00005	19->22	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00005	184->187	PKC_PHOSPHO_SITE	PDOC00005
PS00005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	253->256	PKC_PHOSPHO_SITE	PDOC00005
PS00005	266->269	PKC_PHOSPHO_SITE	PDOC00005
PS00005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS00005	583->586	PKC_PHOSPHO_SITE	PDOC00005
PS00005	601->604	PKC_PHOSPHO_SITE	PDOC00005
PS00005	604->607	PKC_PHOSPHO_SITE	PDOC00005
PS00005	642->645	PKC_PHOSPHO_SITE	PDOC00005
PS00005	662->665	PKC_PHOSPHO_SITE	PDOC00005
PS00006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	55->59	CK2_PHOSPHO_SITE	PDOC00006
PS00006	85->89	CK2_PHOSPHO_SITE	PDOC00006
PS00006	93->97	CK2_PHOSPHO_SITE	PDOC00006
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	168->172	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	266->270	CK2_PHOSPHO_SITE	PDOC00006
PS00006	294->298	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00006	326->330	CK2_PHOSPHO_SITE	PDOC00006
PS00006	337->341	CK2_PHOSPHO_SITE	PDOC00006

PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00006	389->393	CK2_PHOSPHO_SITE	PDOC00006
PS00006	467->471	CK2_PHOSPHO_SITE	PDOC00006
PS00006	514->518	CK2_PHOSPHO_SITE	PDOC00006
PS00006	543->547	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	583->587	CK2_PHOSPHO_SITE	PDOC00006
PS00006	617->621	CK2_PHOSPHO_SITE	PDOC00006
PS00006	658->662	CK2_PHOSPHO_SITE	PDOC00006
PS00006	686->690	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00006	709->713	CK2_PHOSPHO_SITE	PDOC00006
PS00006	714->718	CK2_PHOSPHO_SITE	PDOC00006
PS00006	741->745	CK2_PHOSPHO_SITE	PDOC00006
PS00007	223->230	TYR_PHOSPHO_SITE	PDOC00007
PS00007	222->230	TYR_PHOSPHO_SITE	PDOC00007
PS00008	239->245	MYRISTYL	PDOC00008
PS00008	427->433	MYRISTYL	PDOC00008
PS00008	502->508	MYRISTYL	PDOC00008
PS00008	539->545	MYRISTYL	PDOC00008
PS00008	548->554	MYRISTYL	PDOC00008
PS00008	627->633	MYRISTYL	PDOC00008
PS00009	220->224	AMIDATION	PDOC00009
PS00009	662->666	AMIDATION	PDOC00009
PS00478	390->425	LIM_DOMAIN_1	PDOC00382

Pfam for DKFZphutel_18i19.1

```
HMM_NAME    LIM domain containing proteins

HMM             *CagCNrpIyDREivMRAMNKvWHpECFrCcdCqqPLtegdeFYErDGrI
                 C  C++++Y+ E++  A+  V+H++CFRC+ C+  L+ G+ + ++ GRI
Query       390  CVECQKTVYPMERLL-ANQQVFHISCFRCSYCNNKLSLGT-YASLHGRI    436

HMM             YCKhDYYrrFg*
                YCK+++ ++F+
Query       437 YCKPHFNQLFK    447
```

DKFZphutel_18i4

----

group: uterus derived

DKFZphutel_18i4 encodes a novel 220 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

weak similarity to C.elegans D2085.2

complete cDNA, complete cds, few EST hits

Sequenced by AGOWA

Locus: /map="7q31"

Insert length: 1568 bp
Poly A stretch at pos. 1551, polyadenylation signal at pos. 1523

```
   1 GCCGAGCGGA GAGGGTAGAG ACGGGGTTTC ACCGTGTTAG CCAAGATGGT
  51 CTCGATCTCC TGACCTCGTG ATCCGCCCGC CTCGGCCTCC CAAAGTGCTG
 101 GGATTACAGG CGTGAGCCAC TGCGCCCGGC CTGTTGTACA GTTATTAAAG
 151 TTATCATTTA ACATGGAAGA AGATGAGTTC ATTGGAGAAA AAACATTCCA
 201 ACGTTATTGT GCAGAATTCA TTAAACATTC ACAACAGATA GGTGATAGTT
 251 GGGAATGGAG ACCATCAAAG GACTGTTCTG ATGGCTACAT GTGCAAAATA
 301 CACTTTCAAA TTAAGAATGG GTCTGTGATG TCACATCTAG GAGCATCTAC
 351 CCATGGACAG ACATGTCTTC CCATGGAGGA GGCTTTCGAG CTACCCTTGG
 401 ATGATTGTGA AGTGATTGAA ACTGCAGCAG CGTCCGAAGT GATTAAATAT
 451 GAGTATCATG TCTTATATTC CTGTAGCTAC CAAGTGCCTG TACTTTACTT
 501 TAGGGCAAGC TTTTTAGATG GGAGACCTTT AACTCTGAAG GACATATGGG
 551 AAGGAGTTCA TGAGTGCTAT AAGATGCGAC TGCTACAGGG ACCATGGGAC
 601 ACTATTACGC AACAGGAACA TCCAATACTT GGGCAACCCT TTTTTGTACT
 651 TCATCCCTGC AAGACGAATG AATTCATGAC TCCTGTATTA AAGAATTCTC
 701 AGAAAATCAA TAAGAATGTC AACTATATCA CATCATGGCT GAGCATTGTA
 751 GGGCCAGTTG TTGGGCTGAA TCTACCTCTG AGTTATGCCA AAGCAACGTC
 801 TCAGGATGAA CGAAATGTCC CTTAACAAGA TTCTTCTATT GAGTTTAGGA
 851 ATTGCGGCAC GAAGAATGCC AAGAGTTTAC CTGGCCAGCC CTGGCTTTAA
 901 TAGGACTGAT ACCATGGAAT ATTTCATCTC ACCAAGATGT GACATGGATT
 951 ATTTTTCCCT TGGACACAAA TGTCTACAGC AACTGATGTT TGATAGGCTG
1001 AATGTTTAGA AGAAACACTT CAAAGGGATA CATCATGGCC AGGCATGGTG
1051 GCTCACACCT GTAATCCAAG CACTTTGGGA GGCCAAGGTG GGAGCATCAC
1101 TTGATCCTGG GAGTTCGAGA CCAGCCTGGG CAACATGGTG AAACCCTGTC
1151 GGTACAAAAA AATACAAAAA TTTGCCTGTT TATGGTGGTG TGTTCCTGTA
1201 GTCCCAGCTC CCCAGGAGGC TGAGGTGGGA GGTTGGCTTT AACCCAGGAG
1251 GCAGAGGTTG CAGTGAGCTG AGACTGTGCC ACTGCAGTCC AGCCTGGGTG
1301 ACAGAGCCAG ACACTGTCTC GGGAAAAAAA AAAAAAAAAA AAAGACACAT
1351 CACTATAAAT AGCAAAAAAA CAAATCTAAC TTATTAATAC TAGGAATACC
1401 AACATTATTA GGGCACTTGC AGGTTATTCT TTTCTAGGCC AAGTACTTCA
1451 CTTCCATTTG TCTGACATGG AGATTGAGGG AGAAATGTAT TTGTGTGTTC
1501 ATTTTAATGT AAGATATATA AAAATTAAAT TACTGGATTT ACCTGTCCCT
1551 GAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 163 bp to 822 bp; peptide length: 220
Category: similarity to unknown protein

```
  1 MEEDEFIGEK TFQRYCAEFI KHSQQIGDSW EWRPSKDCSD GYMCKIHFQI
 51 KNGSVMSHLG ASTHGQTCLP MEEAFELPLD DCEVIETAAA SEVIKYEYHV
101 LYSCSYQVPV LYFRASFLDG RPLTLKDIWE GVHECYKMRL LQGPWDTITQ
151 QEHPILGQPF FVLHPCKTNE FMTPVLKNSQ KINKNVNYIT SWLSIVGPVV
201 GLNLPLSYAK ATSQDERNVP
```

BLASTP hits

Entry CED2085_2 from database TREMBL:
"D2085.2"; Caenorhabditis elegans cosmid D2085
Length = 173
Score = 167 (58.8 bits), Expect = 1.1e-12, P = 1.1e-12
Identities = 36/121 (29%), Positives = 64/121 (52%)

Alert BLASTP hits for DKFZphutel_18i4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphutel_18i4, frame 1

Report for DKFZphutel_18i4.1

[LENGTH]	220	
[MW]	25278.99	
[pI]	5.34	
[HOMOL]	TREMBL:CED2085_2 gene: "D2085.2"; Caenorhabditis elegans cosmid D2085 2e-11	
[BLOCKS]	BL00221E	
[PROSITE]	MYRISTYL	2
[PROSITE]	CK2_PHOSPHO_SITE	4
[PROSITE]	PKC_PHOSPHO_SITE	2
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	Alpha_Beta	

```
SEQ  MEEDEFIGEKTFQRYCAEFIKHSQQIGDSWEWRPSKDCSDGYMCKIHFQIKNGSVMSHLG
PRD  cccccccchhhhhhhhhhhhhhhcccccccccccccccceeeeeeeeeeeeccceeeeec

SEQ  ASTHGQTCLPMEEAFELPLDDCEVIETAAASEVIKYEYHVLYSCSYQVPVLYFRASFLDG
PRD  cccccccchhhhhhhhcccccceeehhhhhchhhhhhhheeeecccccceeeeeeecccccc

SEQ  RPLTLKDIWEGVHECYKMRLLQGPWDTITQQEHPILGQPFFVLHPCKTNEFMTPVLKNSQ
PRD  ccccchhhhhhhhhhhhhhhcccccccccccccccccceeeeccccccccccccccccccc

SEQ  KINKNVNYITSWLSIVGPVVGLNLPLSYAKATSQDERNVP
PRD  cccccccccccceeeeccccccccccceeeecccccccccc
```

Prosite for DKFZphutel_18i4.1

PS00001	52->56	ASN_GLYCOSYLATION	PDOC00001
PS00005	124->127	PKC_PHOSPHO_SITE	PDOC00005
PS00005	179->182	PKC_PHOSPHO_SITE	PDOC00005
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	124->128	CK2_PHOSPHO_SITE	PDOC00006
PS00006	149->153	CK2_PHOSPHO_SITE	PDOC00006
PS00006	212->216	CK2_PHOSPHO_SITE	PDOC00006
PS00008	53->59	MYRISTYL	PDOC00008
PS00008	131->137	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphutel_18i4.1)

DKFZphutel_1811
----

group: nucleic acid management

DKFZphtes3_15j18 encodes a novel 184 amino acid protein with similarity to S. cerevisiae putative ribosomal protein YHR148w.

The novel protein is similar to several 40S ribosomal proteins and therefore seems to part of the corresponding ribosome subunit.

The new protein can find application in modulation of ribosome assembly, structure and function.

strong similarity to S.cerevisiae YHR148w

complete cDNA, complete cds, EST hits,
potential start at Bp 45 matchs kozak consensus ANNatgG
gene disruption of YHR148w is lethal!

Sequenced by AGOWA

Locus: unknown

Insert length: 1076 bp
Poly A stretch at pos. 1035, polyadenylation signal at pos. 1006

```
   1 GCGCGCTCTC AGCTTCGGGT CCTGCGGCTG CGGCTGCCGC CATCATGGTG
  51 CGGAAGCTTA AGTTCCACGA GCAGAAGCTG CTGAAGCAGG TGGACTTCCT
 101 GAACTGGGAG GTCACCGACC ACAACCTGCA CGAGCTGCGC GTGCTGCGGC
 151 GTTACCGGCT GCAGCGGCGG GAGGACTACA CGCGCTACAA CCAGCTGAGC
 201 CGTGCCGTGC GTGAGCTGGC GCGGCGCCTG CGCGACCTGC CCGAACGCGA
 251 CCAGTTCCGC GTGCGCGCTT CGGCCGCGCT GCTGGACAAG CTGTATGCTC
 301 TCGGCTTGGT GCCCACGCGC GGTTCGCTGG AGCTCTGCGA CTTCGTCACG
 351 GCCTCGTCCT TCTGCCGCCG CCGCCTCCCC ACCGTGCTCC TCAAGCTGCG
 401 CATGGCGCAG CACCTTCAGG CTGCCGTGGC CTTTGTGGAG CAAGGGCACG
 451 TACGCGTGGG CCCTGACGTG GTTACCGACC CCGCCTTCCT TGTCACGCGC
 501 AGCATGGAGG ACTTTGTCAC TTGGGTGGAC TCGTCCAAGA TCAAGCGGCA
 551 CGTGCTAGAG TACAATGAGG AGCGCGATGA CTTCGATCTG GAAGCCTAGC
 601 GGATCTCCCA CTTTGCATGG CTGTCTTTTA CAGATGGGAA AACTGAGGCC
 651 TGATGCTGGA GATTCTATGA GGGTGCTCTC CTCAAGGGTA TCAGACGGTC
 701 GTAGGTTCTT AAGAATTTGA TTCATCAGTG GCAGGCCATG CATAGAGCCA
 751 CGGGAGGTGC GTCCTTGTTT TCCAGGAAAT GTTCTTAGAA CTTGGACTAC
 801 TGATTATTAA TTGACTGTGC CTTGGGAAAC AGTGGGAAGT AACTTGGTGC
 851 AGCACTGGGG TATTGTTGGA CTGGTTCAAT TCGTTTAACT CGAATTCTTG
 901 CTCCTGGCCG TGGTTAAGCT GTGTACAGAT GATGGAGAGT TTGGCCTCAA
 951 GTTTTTATAA ACTGAGCGAG ACTAGTGTTC AGGATCTCCT CCCTTGTTTA
1001 AATGTCAATA AATGCCCCAA CTGCTTTGTA AGCTCAAAAA AAAAAAAAAA
1051 AAAAAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 45 bp to 596 bp; peptide length: 184
Category: strong similarity to known protein

```
  1 MVRKLKFHEQ KLLKQVDFLN WEVTDHNLHE LRVLRRYRLQ RREDYTRYNQ
 51 LSRAVRELAR RLRDLPERDQ FRVRASAALL DKLYALGLVP TRGSLELCDF
101 VTASSFCRRR LPTVLLKLRM AQHLQAAVAF VEQGHVRVGP DVVTDPAFLV
151 TRSMEDFVTW VDSSKIKRHV LEYNEERDDF DLEA
```

BLASTP hits

```
No BLASTP hits available

             Alert BLASTP hits for DKFZphutel_1811, frame 3

No Alert BLASTP hits found

             Pedant information for DKFZphutel_1811, frame 3
             ----------------------------------------------

                       Report for DKFZphutel_1811.3

[LENGTH]        184
[MW]            21850.21
[pI]            9.54
[HOMOL]         PIR:S33911 probable ribosomal protein YHR148w - yeast (Saccharomyces
cerevisiae) 4e-47
[FUNCAT]        05.01 ribosomal proteins      [S. cerevisiae, YHR148w] 2e-48
[FUNCAT]        30.03 organization of cytoplasm       [S. cerevisiae, YPL081w] 5e-07
[FUNCAT]        j mrna translation and ribosome biogenesis    [M. jannaschii, MJ0190] 8e-05
[BLOCKS]        BL00632
[PIRKW]         cytosol 1e-07
[PIRKW]         ribosome 1e-07
[PIRKW]         protein biosynthesis 1e-07
[SUPFAM]        rat ribosomal protein S9 1e-07
[PROSITE]       MYRISTYL        1
[PROSITE]       CK2_PHOSPHO_SITE        2
[PROSITE]       TYR_PHOSPHO_SITE        1
[PROSITE]       PKC_PHOSPHO_SITE        1
[PFAM]          Ribosomal protein S4
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY      6.52 %

SEQ     MVRKLKFHEQKLLKQVDFLNWEVTDHNLHELRVLRRYRLQRREDYTRYNQLSRAVRELAR
SEG     ................................xxxxxxxxxxxx..................
PRD     ccchhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ     RLRDLPERDQFRVRASAALLDKLYALGLVPTRGSLELCDFVTASSFCRRRLPTVLLKLRM
SEG     ............................................................
PRD     hhhhhccccchhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ     AQHLQAAVAFVEQGHVRVGPDVVTDPAFLVTRSMEDFVTWVDSSKIKRHVLEYNEERDDF
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhcccccceeeccccceeeeecccccceeeeecccchhhhhhhhhcccccc

SEQ     DLEA
SEG     ....
PRD     cccc

                      Prosite for DKFZphutel_1811.3

PS00005     163->166    PKC_PHOSPHO_SITE         PDOC00005
PS00006     153->157    CK2_PHOSPHO_SITE         PDOC00006
PS00006     159->163    CK2_PHOSPHO_SITE         PDOC00006
PS00007      41->49     TYR_PHOSPHO_SITE         PDOC00007
PS00008      87->93     MYRISTYL                 PDOC00008

                       Pfam for DKFZphutel_1811.3

HMM_NAME        Ribosomal protein S4

HMM                 *MSR.YRGPRWKIIRRPGElPWLTnK.....tklmrkYC..lRPgQHgWR
                     M+R ++ +++K+++++++L W          ++++R Y    R+++  ++
Query            1  MVRKLKFHEQKLLKQVDFLNWEVTDHNLHELRVLRRYRLQRREDYTRYN    49

HMM                 qRktLsKIRRmSQYrIRLQEKQKLRFMYGNItERQLRRYvRiaEdKRKlD
                    Q +     +R +++ +   L+E + +R      +++++L++++     +++ L
Query           50  QLSR--AVRELARRLRDLPERDQFRVRASAALLDKLYALGLVP-TRGSLE    96

HMM                 YsTGenLMQILEMRLDNIVFRMGMAPTIHHARQLINHRHIRVNdRIVNIP
                      ++ + ++++RL++++ ++ MA    ++A+ +++++H+RV++ +V++P
Query           97  LCDFVTASSFCRRRLPTVLLKLRMAQHLQAAVAFVEQGHVRVGPDVVTDP   146

HMM                 SYiCRPNDiISIRDkqrMQsHIkWnieSPegrmRPNHLErNnkkYeGtIN
```

```
                 ++++++     +        +++++W++ S+        ++R+ + Y+    +
Query        147 AFLVTRS---M-------EDFVTWVDSSK--------IKRHVLEYNEERD   178

HMM              rIIEReWiplkINElLVVEY*
                 +++ +
Query        179 DFDLE---------------      183
```

DKFZphutel_19f19
----

group: transmembrane protein

DKFZphutel_19f19 encodes a novel 204 amino acid protein with similarity to murine p24 protein.

Murine p24 is expressed only in brain where it is localized exclusively in neurons. It seems to be a neuron-specific membrane protein localised in intracellular organelles of highly differentiated neural cells and may play a role in the neural organelle transport system. As p24, the novel protein contains 2 transmembrane regions, but it contains not the sequence homologous to the microtubule-binding domain of microtubule-associated proteins present in p24.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

similarity to mouse P24 protein ;
membrane regions: 2
Summary DKFZphutel_19f19 encodes a novel 204 amino acid protein, with similarity to mouse P24 protein.

similarity to mouse P24 protein

complete cDNA, complete cds, EST hits,
2 TM-domains

Sequenced by AGOWA

Locus: /map=14.8 cR from top of Chr20 linkage group

Insert length: 2042 bp
Poly A stretch at pos. 1958, polyadenylation signal at pos. 1940

```
   1 GCAGGCAGAG AGATGAGGAA ACTGAGACCC AGAAAGGTGG AAGCACTTGT
  51 CTAAGGTCAC GCCTCCAGGA AGCAGTGTGT CCACGACTCC AGTCCAAGTG
 101 GTCAGGCTCC AGAGCCCACA GTCCCAGGGG TCCATGATGC CGAGCTGCAA
 151 TCGTTCCTGC AGCTGCAGCC GCGGCCCCAG CGTGGAGGAT GGCAAGTGGT
 201 ATGGGGTCCG CTCCTACCTG CACCTCTTCT ATGAGGACTG TGCAGGCACT
 251 GCTCTCAGCG ACGACCCTGA GGGACCTCCG GTCCTGTGCC CCCGCCGGCC
 301 CTGGCCCTCA CTGTGTTGGA AGATCAGCCT GTCCTCGGGG ACCCTGCTTC
 351 TGCTGCTGGG TGTGGCGGCT CTGACCACTG GCTATGCAGT GCCCCCCAAG
 401 CTGGAGGGCA TCGGTGAGGG TGAGTTCCTG GTGTTGGATC AGCGGGCAGC
 451 CGACTACAAC CAGGCCCTGG GCACCTGTCG CCTGGCAGGC ACAGCGCTCT
 501 GTGTGGCAGC TGGAGTTCTG CTCGCCATCT GCCTCTTCTG GGCCATGATA
 551 GGCTGGCTGA GCCAGGACAC CAAGGCAGAG CCCTTGGACC CCGAAGCCGA
 601 CAGCCACGTG GAGGTCTTCG GGGATGAGCC AGAGCAGCAG TTGTCACCCA
 651 TTTTCCGCAA TGCCAGTGGC CAGTCATGGT TCTCGCCACC CGCCAGCCCC
 701 TTTGGGCAAT CTTCTGTGCA GACTATCCAG CCCAAGAGGG ACTCCTGAGC
 751 TGCCCACATG GCCTAAGATG TGGGTCCTGG ATCCTTCCCC CTTCTCACCA
 801 TAACCCCCTC TCAGTGTTTC CCCAACTTCT CCCTTTAGAG CCCAACTCCA
 851 GGTCAAATCT GGAGCTCAAA TCCCAGTGCT CCCTCCCCAG GAGTGGGGCC
 901 CCAACTCTTC CAAGATACCA GCATTCCTCA AGTCCTCCCA AAACTTCCTA
 951 CCCACACCCT CTTCCCAAGG CCCTCAGGGG CAGAAAACAT CTCCTTCAAC
1001 CCGTCCCCAC TCCTTCCTCT GCATGACCTT GGGCAAACCC TTGCCCTTTC
1051 AAGCCATCAG CTCCTGCCTC TCTGCCATGA GGGCTTTGGA TCAGATTCCT
1101 CTTCTCGCCA GGATGAGGAC ACGCACTGCC CTCCATAGAC ACAGATGAAG
1151 GGGTGGGGGT CATTCAGCTC GAATGGGTCC CAGATGCTCA CTTGGCCTTT
1201 CCCTGCAGGA TGAGTGAAGA CGTTTGCCTC TCACAGTGTG TCTTCTACCT
1251 GCATTTTGGC ATCAGAGCCC CCCAGCCCAC CCACCACAGG CAATTACTAG
1301 CCCTAGTTGA TAGGTGAGGT GGGTGAAGAA GGCTGGAGGT GACATGTCCG
1351 AGGTCACACA ACAAAGCAGC ATGCAGGAAC TAGAAACACA TCTTCAGCCT
1401 CCTCCTGGGC CAGCTCTTGT GCTACAGGTG GGGCGGAGCC AGCCCCTCAC
1451 CTTCCTGGTT CCCTGAGGGT CCTCAGGGTG GAGGACAGGT TTGGCCCAGA
1501 AAGACTAGCC AGAGGCCTGA TGGTCCCAGG TGGCTCTGGA TATACTTTGG
1551 ATATGGATTT AAATGGTCTC TAAGAGCCGG GGGTAGGGGG CAGGAAAAGT
1601 GGGTTGTCTT TGCCCCTCAA AGTCCACCTA CCTAGAAACC AAGCCCACGG
1651 TCTTGGCCGT GACCCTGATA ATAAATGGGC TCTCTCAGAG GCGCCAGCCC
1701 CTCCCTCCCC AGCCGGAGGC GTCATCTCTC TTCTGTACCA CTAGAGGGAG
1751 CTCTGATGCA GCTGGAGAGC AGCGCTCAAG GCTCTCGCCC CTCCCCTCCC
1801 TAACCCTTAC CTTCAGTCTC CACCAGCCTG AAGGGCCTCC TAGGGGATCC
1851 TCAGGCGGCC CCCACCAGGG CACACCCTAC TGTCCTTGTG CCTCACGCCC
1901 CCTCCTCATC CTGCACCCCT TCCATCCCAC CTTCCCTTTC AATAAACAGC
1951 TGGGATGGAA AAAAAAAAAA AGAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2001 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA
```

BLAST Results

Entry HS417348 from database EMBL:
human STS WI-14697.
Length = 290
Minus Strand HSPs:
Score = 1254 (188.2 bits), Expect = 3.0e-50, P = 3.0e-50
Identities = 262/273 (95%)

Medline entries

97334404:
A newly identified membrane protein localized exclusively in intracellular organelles of neurons.

Peptide information for frame 2

ORF from 134 bp to 745 bp; peptide length: 204
Category: similarity to known protein

```
  1 MMPSCNRSCS CSRGPSVEDG KWYGVRSYLH LFYEDCAGTA LSDDPEGPPV
 51 LCPRRPWPSL CWKISLSSGT LLLLLGVAAL TTGYAVPPKL EGIGEGEFLV
101 LDQRAADYNQ ALGTCRLAGT ALCVAAGVLL AICLFWAMIG WLSQDTKAEP
151 LDPEADSHVE VFGDEPEQQL SPIFRNASGQ SWFSPPASPF GQSSVQTIQP
201 KRDS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_19f19, frame 2

TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein, complete cds., N = 1, Score = 295, P = 3.8e-26

```
>TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein,
     complete cds.
          Length = 196

  HSPs:

 Score = 295 (44.3 bits), Expect = 3.8e-26, P = 3.8e-26
 Identities = 58/139 (41%), Positives = 81/139 (58%)

Query:     2 MPSCNRSCSCSRGPSVEDGKW---YGVRSYLHLFYEDCAGTALSDDPEGPPVLCPRRPWP 58
             M SC+ +C   R  +  +G +   YGVRSYLH FYEDC  +     + +      P R W
Sbjct:     1 MTSCSNTCGSRRAQADTEGGYQQRYGVRSYLHQFYEDCTASIWEYEDDFQIQRSPNR-WS 59

Query:    59 SLCWKISLSSGTLLLLLGVAALTTGYAVPPKLEGIGEGEFLVLDQRAADYNQALGTCRLA 118
             S+ WK+ L SGT+ ++LG+  L  G+ VPPK+E  GE +F+V+D  A  YN AL TC+LA
Sbjct:    60 SVFWKVGLISGTVFVILGLTVLAVGFLVPPKIEAFGEADFMVVDTHAVKYNGALDTCKLA 119

Query:   119 GTALCVAAGVLLAICLFWAM 138
             G  L    G  +A CL  ++
Sbjct:   120 GAVLFCIGGTSMAGCLLMSV 139
```

Pedant information for DKFZphutel_19f19, frame 2

Report for DKFZphutel_19f19.2

[LENGTH]	204
[MW]	21983.07
[pI]	4.69
[HOMOL]	TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein, complete cds. 7e-19
[PROSITE]	MYRISTYL 4

```
[PROSITE]       CAMP_PHOSPHO_SITE       1
[PROSITE]       CK2_PHOSPHO_SITE        3
[PROSITE]       PKC_PHOSPHO_SITE        1
[PROSITE]       ASN_GLYCOSYLATION       2
[KW]            TRANSMEMBRANE  2
[KW]            LOW_COMPLEXITY     10.29 %

SEQ    MMPSCNRSCSCSRGPSVEDGKWYGVRSYLHLFYEDCAGTALSDDPEGPPVLCPRRPWPSL
SEG    ............................................................
PRD    cccccccccccccccccccccceeehhhhhcccccccccccccccccccccccccccccce
MEM    ..........................................................MM

SEQ    CWKISLSSGTLLLLLGVAALTTGYAVPPKLEGIGEGEFLVLDQRAADYNQALGTCRLAGT
SEG    ....xxxxxxxxxxxxxxxxxxx.....................................
PRD    eeeeeccccceeeccccceeeeccccccccccccccccceeeeccccccccchhhhhhhchh
MEM    MMMMMMMMMMMMMMMMMMMMMMMMMMM.........................MMMMMMM

SEQ    ALCVAAGVLLAICLFWAMIGWLSQDTKAEPLDPEADSHVEVFGDEPEQQLSPIFRNASGQ
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhhhcccccccccccccccceeeecccccccccccccccccccc
MEM    MMMMMMMMMMMMMMMMMMMMMM......................................

SEQ    SWFSPPASPFGQSSVQTIQPKRDS
SEG    ........................
PRD    ccccccccccccceeeeccccccc
MEM    ........................
```

Prosite for DKFZphutel_19f19.2

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	176->180	ASN_GLYCOSYLATION	PDOC00001
PS00004	201->205	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	114->117	PKC_PHOSPHO_SITE	PDOC00005
PS00006	16->20	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	157->161	CK2_PHOSPHO_SITE	PDOC00006
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	92->98	MYRISTYL	PDOC00008
PS00008	119->125	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphutel_19f19.2)

DKFZphutel_19g19
----

group: uterus derived

DKFZphutel_19g19 encodes a novel 400 amino acid protein, with strong but partial similarity to a bovine elastin-related protein expressed in fetal calf ligamentum nuchae.

The novel protein contains 2 RGD cell attachment sites.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

similarity to bovine elastin fragment

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: map=54.9 cR from top of Chr3 linkage group

Insert length: 3244 bp
Poly A stretch at pos. 3227, polyadenylation signal at pos. 3216

```
   1 GTAACTGCAG TAAGTCCCGC TTGGCCCTGG AGTCCACGCG GATTTTCGAA
  51 GCTGGGGCTG GCAAGAGGCC GCTGGACACC ACGCTCCAGT CGTCAGCCCA
 101 CTTCCTAGCT GAACAGCGCG AGGCGGCGGC AGCGAGCCGG GTCCCACCAT
 151 GGCCGCGAAT TATTCCAGTA CCAGTACCCG GAGAGAACAT GTCAAAGTTA
 201 AAACCAGCTC CCAGCCAGGC TTCCTGGAAC GGCTGAGCGA GACCTCGGGT
 251 GGGATGTTTG TGGGGCTCAT GGCCTTCCTG CTCTCCTTCT ACCTAATTTT
 301 CACCAATGAG GGCCGCGCAT TGAAGACGGC AACCTCATTG GCTGAGGGGC
 351 TCTCGCTTGT GGTGTCTCCT GACAGCATCC ACAGTGTGGC TCCGGAGAAT
 401 GAAGGAAGGC TGGTGCACAT CATTGGCGCC TTACGGACAT CCAAGCTTTT
 451 GTCTGATCCA AACTATGGGG TCCATCTTCC GGCTGTGAAA CTGCGGAGGC
 501 ACGTGGAGAT GTACCAATGG GTAGAAACTG AGGAGTCCAG GGAGTACACC
 551 GAGGATGGGC AGGTGAAGAA GGAGACGAGG TATTCCTACA ACACTGAATG
 601 GAGGTCAGAA ATCATCAACA GCAAAAACTT CGACCGAGAG ATTGGCCACA
 651 ATAACCCCAG TGCCATGGCA GTGGAGTCAT TCACGGCAAC AGCCCCCTTT
 701 GTCCAAATTG GCAGGTTTTT CCTCTCGTCA GGCCTCATCG ACAAAGTCGA
 751 CAACTTCAAG TCCCTGAGCC TATCCAAGCT GGAGGACCCT CATGTGGACA
 801 TCATTCGCCG TGGAGACTTT TTCTACCACA GCGAAAATCC CAAGTATCCA
 851 GAGGTGGGAG ACTTGCGTGT CTCCTTTTCC TATGCTGGAC TGAGCGGCGA
 901 TGACCCTGAC CTGGGCCCAG CTCACGTGGT CACTGTGATT GCCCGGCAGC
 951 GGGGTGACCA GCTAGTCCCA TTCTCCACCA AGTCTGGGGA TACCTTACTG
1001 CTCCTGCACC ACGGGGACTT CTCAGCAGAG GAGGTGTTTC ATAGAGAACT
1051 AAGGAGCAAC TCCATGAAGA CCTGGGGCCT GCGGGCAGCT GGCTGGATGG
1101 CCATGTTCAT GGGCCTCAAC CTTATGACAC GGATCCTCTA CACCTTGGTG
1151 GACTGGTTTC CTGTTTTCCG AGACCTGGTC AACATTGGCC TGAAAGCCTT
1201 TGCCTTCTGT GTGGCCACCT CGCTGACCCT GCTGACCGTG GCGGCTGGCT
1251 GGCTCTTCTA CCGACCCCTG TGGGCCCTCC TCATTGCCGG CCTGGCCCTT
1301 GTGCCCATCC TTGTTGCTCG GACACGGGTG CCAGCCAAAA AGTTGGAGTG
1351 AAAAGACCCT GGCACCCGCC CGACACCTGC GTGAGCCCTA GGATCCAGGT
1401 CCTCTCTCAC CTCTGACCCA GCTCCATGCC AGAGCAGGAG CCCCGGTCAA
1451 TTTTGGACTC TGCACCCCCT CTCCTCTTCA GGGGCCAGAC TTGGCAGCAT
1501 GTGCACCAGG TTGGTGTTCA CCAGCTCATG TCTTCCCCAC ATCTCTTCTT
1551 GCCAGTAAGC AGCTTTGGTG GGCAGCAGCA GCCATGAATG GCAAGCTGAC
1601 AGCTTCTCCT GCTGTTTCCT TCCTCTCTTG GACTGAGTGG GTACGGCCAG
1651 CCACTCAGCC CATTGGCAGC TGACAACGCA GACACGCTCT ACGGAGGCCT
1701 GCTGATAAAG GGCTCAGCCT TGCCGTGTGC TGCTTCTCAT CACTGCACAC
1751 AAGTGCCATG CTTTGCCACC ACCACCAAGC ACATCTGTGA TCCTGAAGGG
1801 CGGCCGTTAG TCATTACTGC TGAGTCCTGG GTCACCAGCA GACACACTGG
1851 GCATGGACCC CTCAAAGCAG GCACACCCAA AACACAAGTC TGTGGCTAGA
1901 ACCTGATGTG GTGTTTAAAA GAGAAGAAAC ACTGAAGATG TCCTGAGGAG
1951 AAAAGCTGGA CATATACTGG GCTTCACACT TATCTTATGG CTTGGCAGAA
2001 TCTTTGTAGT GTGTGGGATC TCTGAAGGCC CTATTTAAGT TTTTCTTCGT
2051 TACTTTGCTG CTTCATGTGT ACTTTCCTAC CCCAAGAGGA AGTTTTCTGA
2101 AATAAGATTT AAAAACAAAA CAAAAAAAAC ACTTAATATT TCAGACTGTT
2151 ACAGGAAACA CCCTTTAGTC TGTCAGTTGA ATTCAGAGCA CTGAAAGGTG
2201 TTAAATTGGG GTATGTGGTT TGATTGATAA AAAGTTACCT CTCAGTATTT
2251 TGTGTCACTG AGAAGCTTTA CAATGGATGC TTTTGAAACA AGTATCAGCA
2301 AAAGGATTTG TTTTCACTCT GGGAGGAGAG GGTGGAGAAA GCACTTGCTT
2351 TCATCCTCTG GCATCGGAAA CTCCCCTATG CACTTGAAGA TGGTTTAAAA
2401 GATTAAAGAA ACGATTAAGA GAAAAGGTTG GAAGCTTTAT ACTAAATGGG
2451 CTCCTTCATG GTGACGCCCC GTCAACCACA ATCAAGAACT GAGGCCTGAG
2501 GCTGGTTGTA CAATGCCCAC GCCTGCCTGG CTGCTTTCAC CTGGGAGTGC
2551 TTTCGATGTG GGCACCTGGG CTTCCTAGGG CTGCTTCTGA GTGGTTCTTT
2601 CACGTGTTGT GTCCATAGCT TTAGTCTTCC TAAATAAGAT CCACCCACAC
```

```
2651 CTAAGTCACA GAATTTCTAA GTTCCCCAAC TACTCTCACA CCCTTTTAAA
2701 GATAAAGTAT GTTGTAACCA GGATGTCTTA AATGATTCTT TGTGTACCTT
2751 TTCTGTCATA TTCAGAAACC GTTTTGTGCC TGCTGGGAGT AATTCCTTTA
2801 GCAATTAAGT ATTTGGTAGC TGAATAAGGG GTCAGAACTT CTGAAACCAG
2851 AGATCTGTAA TCATCTCTAT TGGCCTGGGG TGCCTGTGCT ATAAATGAGT
2901 TTCTTCACAT GAAAAACACA GCCAGCCCAA GATGACTTAT CTGGGTTTAG
2951 GATTCAATAG TATTCACTAA CTGCTTATTA CATGAGCAAT TTCATCAAAT
3001 CTCCAAACTC TTAAAGGATG CTTTCGGAAA ACACGCTGTA TACCTAGATG
3051 ATGACTAAAT GCAAAATCCT TGGGCTTTGG TTTTTTTCTA GTAAGGATTT
3101 TAAATAACTG CCGACTTCAA AAGTGTTCTT AAAACGAAAG ATAATGTTAA
3151 GAAAAATTTG AAAGCTTTGG AAAACCAAAT TTGTAATATC ATTGTATTTT
3201 TTATTAAAAG TTTTGTAATA AATTTCTAAA AAAAAAAAAA AAAA
```

BLAST Results

Entry HS545355 from database EMBL:
human STS WI-14815.
Length = 436
Minus Strand HSPs:
Score = 2040 (306.1 bits), Expect = 6.2e-86, P = 6.2e-86
Identities = 420/426 (98%)

Entry HS932147 from database EMBL:
human STS WI-8531.
Length = 341
Minus Strand HSPs:
Score = 1705 (255.8 bits), Expect = 4.7e-70, P = 4.7e-70
Identities = 341/341 (100%)

Medline entries

86051793:
Bovine elastin cDNA clones: evidence for the occurrence of a new elastin-related protein in fetal calf ligamentum nuchae.

Peptide information for frame 2

ORF from 149 bp to 1348 bp; peptide length: 400
Category: similarity to known protein

```
  1 MAANYSSTST RREHVKVKTS SQPGFLERLS ETSGGMFVGL MAFLLSFYLI
 51 FTNEGRALKT ATSLAEGLSL VVSPDSIHSV APENEGRLVH IIGALRTSKL
101 LSDPNYGVHL PAVKLRRHVE MYQWVETEES REYTEDGQVK KETRYSYNTE
151 WRSEIINSKN FDREIGHNNP SAMAVESFTA TAPFVQIGRF FLSSGLIDKV
201 DNFKSLSLSK LEDPHVDIIR RGDFFYHSEN PKYPEVGDLR VSFSYAGLSG
251 DDPDLGPAHV VTVIARQRGD QLVPFSTKSG DTLLLLHHGD FSAEEVFHRE
301 LRSNSMKTWG LRAAGWMAMF MGLNLMTRIL YTLVDWFPVF RDLVNIGLKA
351 FAFCVATSLT LLTVAAGWLF YRPLWALLIA GLALVPILVA RTRVPAKKLE
```

BLASTP hits

Entry I45887 from database PIR:
elastin - bovine (fragment)
Length = 40
Score = 131 (46.1 bits), Expect = 4.9e-08, P = 4.9e-08
Identities = 31/41 (75%), Positives = 34/41 (82%)

Alert BLASTP hits for DKFZphutel_19g19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphutel_19g19, frame 2

Report for DKFZphutel_19g19.2

[LENGTH] 400

[MW]	44831.53
[pI]	7.23
[HOMOL]	PIR:I45887 elastin - bovine (fragment) 1e-06
[PROSITE]	RGD 2
[PROSITE]	MYRISTYL 3
[PROSITE]	CAMP_PHOSPHO_SITE 1
[PROSITE]	CK2_PHOSPHO_SITE 6
[PROSITE]	TYR_PHOSPHO_SITE 2
[PROSITE]	PKC_PHOSPHO_SITE 5
[PROSITE]	ASN_GLYCOSYLATION 1
[KW]	TRANSMEMBRANE 4

```
SEQ  MAANYSSTSTRREHVKVKTSSQPGFLERLSETSGGMFVGLMAFLLSFYLIFTNEGRALKT
PRD  ccceeeccccceeeeeeeeecccccceeeecccccccchhhhhhhhhheeeeeccccchhhh
MEM  ............................MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM..

SEQ  ATSLAEGLSLVVSPDSIHSVAPENEGRLVHIIGALRTSKLLSDPNYGVHLPAVKLRRHVE
PRD  hhhhhccceeeecccccceeeecccccceeeeeeeeeeceeecccccccccccchhhhhhhhh
MEM  ............................................................

SEQ  MYQWVETEESREYTEDGQVKKETRYSYNTEWRSEIINSKNFDREIGHNNPSAMAVESFTA
PRD  hheeehhhhheeeccccccccceeeccccccceeeeeecccccceeeccccccceeeeeeecc
MEM  ...........................................................M

SEQ  TAPFVQIGRFFLSSGLIDKVDNFKSLSLSKLEDPHVDIIRRGDFFYHSENPKYPEVGDLR
PRD  ccceeeeeeeeeccccccccccceeeeeeeccccceeeeeccccceeecccccccccccccee
MEM  MMMMMMMMMMMMMMMMMM..........................................

SEQ  VSFSYAGLSGDDPDLGPAHVVTVIARQRGDQLVPFSTKSGDTLLLHHGDFSAEEVFHRE
PRD  eeccccccccccccccceeeeeeeeecccccccccccccccceeeeeeccccchhhhhhh
MEM  ............................................................

SEQ  LRSNSMKTWGLRAAGWMAMFMGLNLMTRILYTLVDWFPVFRDLVNIGLKAFAFCVATSLT
PRD  hhccccccccchhhhhhhhhhhchhhhhhhhheeecccccccccccceeeeeeeeehhhh
MEM  ........MMMMMMMMMMMMMMMMMMMMMMMMM.......................MMMM

SEQ  LLTVAAGWLFYRPLWALLIAGLALVPILVARTRVPAKKLE
PRD  hhhhhccceeehhhhhhhhhhhchhhhhhhhcccccccc
MEM  MMMMMMMMMMMMMMMMMMMMMMMMMM...............
```

Prosite for DKFZphutel_19g19.2

PS00001	4->8	ASN_GLYCOSYLATION	PDOC00001
PS00004	140->144	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS00005	10->13	PKC_PHOSPHO_SITE	PDOC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDOC00005
PS00005	276->279	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00006	10->14	CK2_PHOSPHO_SITE	PDOC00006
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00006	249->253	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	332->336	CK2_PHOSPHO_SITE	PDOC00006
PS00007	220->227	TYR_PHOSPHO_SITE	PDOC00007
PS00007	99->107	TYR_PHOSPHO_SITE	PDOC00007
PS00008	35->41	MYRISTYL	PDOC00008
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	310->316	MYRISTYL	PDOC00008
PS00016	221->224	RGD	PDOC00016
PS00016	268->271	RGD	PDOC00016

(No Pfam data available for DKFZphutel_19g19.2)

DKFZphutel_19g22
----

group: cell structure and motility

DKFZphutel_19g22 encodes a novel 390 amino acid protein with very strong similarity to tuftelin/enamelin.

Tuftelin/enamelin are matrix proteins of the teeth. As other proteins involved in calcification, these proteins are also expressed in the uterus matrix.

The new protein can find application in modulation of tissue-calcification, especially the uterus.

complete cDNA, complete cds start at Bp 51, EST hits in 3' UTR,
human homolog of mouse tuftelin
tuftelin is descriebed as a matrix protein of teeth but it seems also
to be pressend in the uterus matrix

Sequenced by AGOWA

Locus: unknown

Insert length: 3110 bp
Poly A stretch at pos. 3093, polyadenylation signal at pos. 3071

```
   1 GCAGACAGCG GGGTGGACAA GTGGCGTGTG TGCTGCGACC CCGAGGGAAG
  51 ATGAACGGGA CGCGGAACTG GTGTACCCTG GTGGACGTGC ACCCAGAGGA
 101 CCAGGCGGCG GGCAGCGTGG ACATTCTCAG GCTGACTCTC CAGGGTGAAC
 151 TGACAGGAGA TGAACTTGAA CACATAGCCC AGAAGGCGGG CAGGAAGACC
 201 TATGCCATGG TGTCCAGCCA CTCAGCTGGT CATTCTCTGG CTTCAGAACT
 251 GGTGGAGTCC CATGATGGAC ATGAGGAGAT CATTAAGGTG TACTTGAAGG
 301 GGAGGTCTGG AGACAAGATG ATTCACGAGA AGAATATTAA CCAGCTGAAG
 351 AGTGAGGTCC AGTACATCCA GGAGGCCAGG AACTGCCTAC AGAAGCTCCG
 401 GGAGGATATA AGTAGCAAGC TTGACAGGAA CCTAGGAGAT TCTCTCCATC
 451 GACAGGAGAT ACAGGTGGTG CTAGAAAAGC CAAATGGCTT TAGTCAGAGT
 501 CCCACAGCCC TGTACAGCAG CCCACCTGAG GTGGACACCT GTATAAATGA
 551 GGATGTTGAG AGCTTGAGGA AGACGGTGCA GGACTTGCTG GCCAAGCTTC
 601 AGGAGGCCAA CGGGCAACAC CAGTCAGACT GTGTGGCTTT TGAGGTCACA
 651 CTCAGCCGGT ACCAGAGGGA AGCAGAACAA AGTAATGTGG CCCTTCAGAG
 701 AGAGGAGGAC AGAGTGGAGC AGAAAGAGGC AGAAGTCGGA GAGCTGCAGA
 751 GGCGCTTGCT AGGGATGGAG ACGGAGCATC AGGCCTTACT GGCGAAAGTG
 801 AGGGAAGGGG AGGTGGCCCT AGAGGAACTT CGGAGCAACA ATGCTGACTG
 851 CCAAGCAGAA CGAGAAAAGG CTGCTACCCT GGAAAAGGAA GTGGCCGGGT
 901 TGCGGGAGAA GATCCACCAC TTGGATGACA TGCTCAAGAG CCAGCAGCGG
 951 AAAGTCCGGC AAATGATAGA GCAGCTCCAG AATTCAAAAG CTGTGATCCA
1001 GTCAAAGGAC GCCACCATCC AGGAGCTCAA GGAGAAAATC GCCTATCTGG
1051 AGGCAGAGAA TTTAGAGATG CATGACCGGA TGGAACACCT GATAGAAAAA
1101 CAAATCAGTC ATGGCAACTT CAGCACCCAG GCCCGGGCCA AGACAGAGAA
1151 CCCGGGCAGT ATTAGGATAT CCAAGCCGCC TAGCCCGAAG CCCATGCCTG
1201 TCATCCGAGT GGTGGAAACC TGAGCTGCCT GGAGATGGTT GCTGCCATTG
1251 CTGCTGCCTC TGCCTCGGAG AAGCCCACTG CCCCTGTTGG CTGTTAACAC
1301 TGCCTTTGAC TTCCTGACTG TCCCCTGGCT GCACCCAGGA CTTCGGGCTC
1351 CTGTGTCTCA CCATTCCCAA GCCCCTGGCC ACTCTAAGCT GGGCAGACGG
1401 AGCACGAGCA CCTATTCAAG GCACTGCAGC CCTTTGGAAG ACATTGTCCT
1451 GCAAGCAGGA GCCAGGGCAA TATCTATATT CCTACAGTGA CTATTTTTCT
1501 CTGTAGAGAG CCTCCCTTCT GTTGTAGACT GGACTCTGGC TGCGCCATAA
1551 GCCAGGCCTT CATCAGATTG GGAGAGGTGA CAAGATTTGC CTCAGCCCTA
1601 AAAGCTGGAG ACACAGATGT CCAGAGTGAT TGGAGAATGT CCTGGGGGAA
1651 TGAAGTTCCT TCCACAAACA CAGCTCAGTT CTTAGCAACA AACTGTTTGT
1701 TTTTCTACTT GCTCCATCTG CAGCCTACGC TGCCCTGGCC TCCTGCAGAC
1751 AGATAGTGGG GTTACCTGGC AAGGCCTGGT GAGAGCCAGT GAACCTAAGC
1801 TTTGACTGGG TGGCCTTGTC TTTCTGGGGA GGAGGGAATG TACATTCAGG
1851 GAGTAGCCTT TTGCGGAAAA ATTCTCTAGG GCTACAGACA GTCATGTGTG
1901 ACTTCTCTCT GCTGTGAAAA CTCCCAGAGT CTCTTTAGGG ATTTTCCCTA
1951 AGGTGTACCA CCAGGCACAC CTCAGTCTTC TTGACCCAGA GCCTGAAAAC
2001 TGTTTTCACT GGGTTCCACC AGTCCCAGCA AAATCCTCTT TGTATTTATT
2051 TTGCTAAGTT ATTGGTGGTT TTGCTTACAT CTCATGATTG ATATAATACC
2101 AAAGTTCTAT AGCCTTCTCT TGCAGTATTT GGATTTGCTT GAAACCGGGA
2151 AAACTGTTCC CATTAGGCTT GTTAATGTCA GAGTGACACT ATTATGAATC
2201 TTTCTCTCCC TTTCCTCTGC CTGTTTCTTC TCTCTTTCTC CTTCAAACTT
2251 GCTCTGCAGC TAAGGAAGGT GAGTCTACTT TCCCTGAGGC TTTGGGGTCA
2301 GAGTATATGT TGTTTGGAGA AAGAGGGCAA TCAGGACTCT TCTGGGACCC
2351 AGATGAGTTC TTCACTAGCC CTTCTGAACC CCTTGCTCCA TAATTGGTCT
2401 TTTATCCTGG CTCTGAATGA CCCTGCAGGT CATCATGGTT TTCTTTTTTT
2451 ATTGTTTTTT TTTTTTTCTG AGACAGAGTC TCACTCTGTC ACCCAGGCTG
2501 GAGTGCAGTG GCGCGATCTC AGCTCACTGC AACCTCTGCC TCCCGGATTT
2551 AAGCGATTCT TCTGCCTCAG CCTCCCGAGT AGCTGGGACT ACAGGTGTGC
```

```
2601 CACCACGCCT GGCTGATTTT TGTATTTTTA GTAGAGATGG GGTTTCACCA
2651 TACTGGCTAG GCTGGTCTCG AATTCCTGAC CTCAGGTGAT CCACCCACCT
2701 CGGCTTCCCA AAGTGCTAGG ATTATAGGCT TGAGCTACTG TGCCCGGCCC
2751 ATGGTGTTTT TCTTTAGGGC TCTTCCTACA GCCTTGAGAA GTAGATAGGC
2801 ATCAGAGTAT GGTACTATAG GAATCAGAAA AATTCAAAAC AAATGTGGAT
2851 TAAGTGTTTA GGCTCTATGT GGCTCACGCA GCCAGAATCC TTAAGTCTGT
2901 GTGTTTCTGT GTCTCAAGAC TGGGCTCACA TTCTGGCTTT GTCCATAACA
2951 ATGCTCTGGG ATTTCAGGGA GTTCCCTCAT TTGTAAAATG AGGGGGTCAG
3001 AGCAGGTGAT ATCCATGTTT CTTCCCTTTC TGATATTGTT GTCTGTGGCA
3051 TATTCTTTGT ATGGCGAATT TAATAAATTA TATTAATGTG TCTAAAAAAA
3101 AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

98200312:
Tuftelin--aspects of protein and gene structure

97228909:
Timing of the expression of enamel gene products during mouse tooth development.

91340750:
Sequencing of bovine enamelin ("tuftelin") a novel acidic enamel protein.

Peptide information for frame 3

ORF from 51 bp to 1220 bp; peptide length: 390
Category: strong similarity to known protein

```
  1 MNGTRNWCTL VDVHPEDQAA GSVDILRLTL QGELTGDELE HIAQKAGRKT
 51 YAMVSSHSAG HSLASELVES HDGHEEIIKV YLKGRSGDKM IHEKNINQLK
101 SEVQYIQEAR NCLQKLREDI SSKLDRNLGD SLHRQEIQVV LEKPNGFSQS
151 PTALYSSPPE VDTCINEDVE SLRKTVQDLL AKLQEAKRQH QSDCVAFEVT
201 LSRYQREAEQ SNVALQREED RVEQKEAEVG ELQRRLLGME TEHQALLAKV
251 REGEVALEEL RSNNADCQAE REKAATLEKE VAGLREKIHH LDDMLKSQQR
301 KVRQMIEQLQ NSKAVIQSKD ATIQELKEKI AYLEAENLEM HDRMEHLIEK
351 QISHGNFSTQ ARAKTENPGS IRISKPPSPK PMPVIRVVET
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_19g22, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphutel_19g22, frame 3

Report for DKFZphutel_19g22.3

```
[LENGTH]      390
[MW]          44264.09
[pI]          5.68
[HOMOL]       TREMBL:AF047704_1 product: "tuftelin"; Mus musculus tuftelin mRNA, complete
cds. 0.0
[FUNCAT]      08.07 vesicular transport (golgi network, etc.)     [S. cerevisiae, YDL058w]
2e-11
[FUNCAT]      30.03 organization of cytoplasm     [S. cerevisiae, YDL058w] 2e-11
[FUNCAT]      1 genome replication, transcription, recombination and repair     [M.
jannaschii, MJ1643] 7e-11
[FUNCAT]      09.13 biogenesis of chromosome structure     [S. cerevisiae, YLR086w] 1e-08
[FUNCAT]      03.22.01 cell cycle check point proteins     [S. cerevisiae, YGL086w] 6e-08
[FUNCAT]      30.10 nuclear organization     [S. cerevisiae, YGL086w] 6e-08
[FUNCAT]      03.13 meiosis   [S. cerevisiae, YNL250w] 7e-08
```

[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 7e-08
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-07
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR285w] 2e-07
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 2e-07
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR216c] 1e-05
[FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YNL243w] 1e-04
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YNL243w] 1e-04
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YNL243w] 1e-04
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YNL243w] 1e-04
[FUNCAT] 08.19 cellular import [S. cerevisiae, YNL243w] 1e-04
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YNL243w] 1e-04
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-04
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-04
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 4e-04
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YMR294w] 7e-04
[EC] 3.6.1.32 Myosin ATPase 8e-09
[PIRKW] blocked amino end 1e-07
[PIRKW] nucleus 1e-06
[PIRKW] citrulline 1e-07
[PIRKW] tandem repeat 8e-09
[PIRKW] heterodimer 3e-06
[PIRKW] DNA repair 2e-06
[PIRKW] heart 8e-09
[PIRKW] endocytosis 3e-07
[PIRKW] transmembrane protein 4e-10
[PIRKW] zinc finger 3e-07
[PIRKW] metal binding 3e-07
[PIRKW] muscle contraction 8e-09
[PIRKW] acetylated amino end 1e-06
[PIRKW] actin binding 8e-09
[PIRKW] microtubule binding 1e-06
[PIRKW] cell division control 1e-06
[PIRKW] ATP 8e-09
[PIRKW] chromosomal protein 3e-06
[PIRKW] thick filament 8e-09
[PIRKW] phosphoprotein 1e-145
[PIRKW] skeletal muscle 8e-09
[PIRKW] calcium binding 1e-07
[PIRKW] meiosis 2e-06
[PIRKW] alternative splicing 7e-08
[PIRKW] DNA condensation 3e-06
[PIRKW] coiled coil 4e-10
[PIRKW] P-loop 8e-09
[PIRKW] heptad repeat 1e-07
[PIRKW] methylated amino acid 8e-09
[PIRKW] immunoglobulin receptor 2e-06
[PIRKW] peripheral membrane protein 3e-07
[PIRKW] cardiac muscle 8e-09
[PIRKW] hydrolase 8e-09
[PIRKW] muscle 7e-08
[PIRKW] EF hand 1e-07
[PIRKW] cytoskeleton 7e-08
[PIRKW] hair 1e-07
[PIRKW] smooth muscle 7e-08
[PIRKW] calmodulin binding 3e-07
[SUPFAM] conserved hypothetical P115 protein 2e-09
[SUPFAM] myosin heavy chain 8e-09
[SUPFAM] RAD50 protein 2e-06
[SUPFAM] calmodulin repeat homology 1e-07
[SUPFAM] myosin motor domain homology 8e-09
[SUPFAM] alpha-actinin actin-binding domain homology 1e-06
[SUPFAM] tropomyosin 7e-08
[SUPFAM] protein-tyrosine kinase ret 3e-07
[SUPFAM] plectin 1e-06
[SUPFAM] trichohyalin 1e-07
[SUPFAM] pleckstrin repeat homology 2e-06
[SUPFAM] ribosomal protein S10 homology 1e-06
[SUPFAM] protein kinase homology 3e-07
[SUPFAM] protein kinase C zinc-binding repeat homology 2e-06
[SUPFAM] giantin 4e-06
[SUPFAM] kinesin-related protein KLPA 1e-06
[SUPFAM] kinesin motor domain homology 1e-06
[SUPFAM] human early endosome antigen 1 3e-07
[SUPFAM] M5 protein 2e-06
[PROSITE] MYRISTYL 1
[PROSITE] AMIDATION 1
[PROSITE] CK2_PHOSPHO_SITE 6

```
[PROSITE]       PKC_PHOSPHO_SITE      4
[PROSITE]       ASN_GLYCOSYLATION     2
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY     4.62 %
[KW]            COILED_COIL       35.13 %

SEQ     MNGTRNWCTLVDVHPEDQAAGSVDILRLTLQGELTGDELEHIAQKAGRKTYAMVSSHSAG
SEG     ............................................................
PRD     ccccccceeeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ............................................................

SEQ     HSLASELVESHDGHEEIIKVYLKGRSGDKMIHEKNINQLKSEVQYIQEARNCLQKLREDI
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ............................................................

SEQ     SSKLDRNLGDSLHRQEIQVVLEKPNGFSQSPTALYSSPPEVDTCINEDVESLRKTVQDLL
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ...........................................CCCCCCCCCCCCCCCCC

SEQ     AKLQEAKRQHQSDCVAFEVTLSRYQREAEQSNVALQREEDRVEQKEAEVGELQRRLLGME
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCCCCCC.....................CCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ     TEHQALLAKVREGEVALEELRSNNADCQAEREKAATLEKEVAGLREKIHHLDDMLKSQQR
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CC....................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ     KVRQMIEQLQNSKAVIQSKDATIQELKEKIAYLEAENLEMHDRMEHLIEKQISHGNFSTQ
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCCCCC........CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC..........

SEQ     ARAKTENPGSIRISKPPSPKPMPVIRVVET
SEG     .........xxxxxxxxxxxxxxxxxx...
PRD     hhccccccccceeeeccccccccccceeeccc
COILS   ..............................
```

Prosite for DKFZphutel_19g22.3

PS00001	2->6	ASN_GLYCOSYLATION	PDOC00001
PS00001	356->360	ASN_GLYCOSYLATION	PDOC00001
PS00005	121->124	PKC_PHOSPHO_SITE	PDOC00005
PS00005	171->174	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	378->381	PKC_PHOSPHO_SITE	PDOC00005
PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	35->39	CK2_PHOSPHO_SITE	PDOC00006
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	157->161	CK2_PHOSPHO_SITE	PDOC00006
PS00006	175->179	CK2_PHOSPHO_SITE	PDOC00006
PS00006	322->326	CK2_PHOSPHO_SITE	PDOC00006
PS00008	355->361	MYRISTYL	PDOC00008
PS00009	46->50	AMIDATION	PDOC00009

(No Pfam data available for DKFZphutel_19g22.3)

DKFZphutel_19h17
----

group: intracellular transport and trafficking

DKFZphutel_19h17 encodes a novel 879 amino acid protein, with similarity to N.crassa osbP oxysterol-binding protein.

The novel protein contains a oxysterol-binding protein family signature. Mammalian oxysterol-binding protein (OSBP) is a protein binds a variety of oxysterols (oxygenated derivatives of cholesterol). OSBP seems to play a complex role in the regulation of sterol metabolism. OSBP is a cytosolic/Golgi receptor for oxysterols such as 25-hydroxycholesterol, and thus a potential target of siphingomyelin turnover and cholesterol mobilization at the plasma membrane and/or Golgi apparatus. Therefore, the new protein seems to be involved in oxysterol metabolism.

The new protein can find application in modulating the response of cells to oxysterols. The protein can be used as marker for the golgi system. The Protein might be used to direct drugs to the golgi system in response to oxidative stess.

strong similarity to C.elegans ZK1086.1 and oxysterol-binding proteins

complete cDNA, complete cds, few EST hits
similarity to proteins involved in steroid biosynthesis

Sequenced by AGOWA

Locus: unknown

Insert length: 3828 bp
Poly A stretch at pos. 3811, polyadenylation signal at pos. 3784

```
   1 GCCCGCGCGC CCGGCCGGCC CGGAGCACCG AGCTCGCGGC ACGGTAGGAG
  51 AAGCCCCCGA GCGCCCACAG CATGAAGGAG GAGGCCTTCC TCCGGCGCCG
 101 CTTCTCCCTG TGTCCACCTT CCTCCACCCC TCAGAAAGTC GACCCCCGGA
 151 AGCTCACCCG GAACTTGCTC CTCAGCGGAG ACAATGAGCT CTACCCACTC
 201 AGCCCAGGGA AGGACATGGA GCCCAACGGC CCGTCGCTGC CCAGGGATGA
 251 AGGGCCCCCG ACCCCAAGCT CTGCCACGAA GGTGCCACCG GCAGAGTACA
 301 GGCTGTGCAA CGGGTCAGAC AAGGAATGTG TGTCCCCCAC CGCCAGGGTC
 351 ACCAAGAAGG AGACTCTCAA GGCGCAGAAG GAGAACTACC GGCAGGAGAA
 401 GAAGCGCGCC ACACGGCAGC TGCTCAGCGC TCTGACAGAC CCCAGCGTGG
 451 TCATCATGGC TGACAGCCTG AAGATCCGCG GCACCCTGAA GAGCTGGACC
 501 AAGCTGTGGT GCGTGCTGAA GCCGGGGGTG CTGCTCATCT ACAAGACGCC
 551 CAAGGTGGGC CAGTGGGTGG GCACGGTGCT GCTGCACTGC TGCGAGCTCA
 601 TCGAGCGGCC CTCCAAGAAG GACGGCTTCT GCTTCAAGCT CTTCCACCCG
 651 CTGGATCAGT CCGTCTGGGC CGTGAAGGGC CCCAAAGGTG AGAGCGTGGG
 701 CTCCATCACA CAGCCCCTGC CCAGCAGCTA CCTGATCTTC AGGGCCGCCT
 751 CCGAGTCAGA TGGTCGCTGC TGGCTGGACG CCCTGGAGCT GGCCCTGCGC
 801 TGCTCTAGCC TACTGAGACT GGGCACCTGC AAGCCGGGCC GAGACGGGGA
 851 GCCAGGGACC TCGCCAGACG CATCACCCTC ATCGCTCTGT GGGCTGCCAG
 901 CCTCAGCCAC TGTCCACCCA GACCAAGACC TGTTCCCACT GAACGGGTCT
 951 TCCCTGGAGA ACGATGCATT CTCAGACAAG TCGGAGAGAG AGAACCCTGA
1001 GGAGTCAGAT ACCGAGACCC AGGACCATAG CCGGAAGACG GAGAGTGGCA
1051 GCGACCAGTC AGAGACCCCT GGGGCCCCGG TGCGGAGAGG GACCACCTAT
1101 GTGGAGCAGG TCCAGGAGGA GCTGGGGGAG CTGGGCGAGG CGTCCCAGGT
1151 GGAGACAGTG TCAGAGGAGA ACAAGAGTCT GATGTGGACC CTGCTGAAGC
1201 AGCTACGGCC AGGCATGGAC CTGTCCCGCG TGGTGCTACC CACGTTCGTA
1251 CTGGAGCCGC GCTCCTTCCT GAACAAGCTC TCCGACTACT ACTACCACGC
1301 AGACCTGCTC TCCAGGGCTG CGGTGGAGGA GGATGCCTAC AGCCGCATGA
1351 AGCTGGTGCT GCGGTGGTAC CTGTCTGGCT TCTACAAGAA GCCCAAGGGA
1401 ATCAAGAAGC CGTACAACCC CATCCTGGGG GAGACCTTCC GCTGCTGCTG
1451 GTTCCACCCG CAGACTGACA GCCGCACATT CTACATAGCA GAGCAGGTGT
1501 CCCACCACCC GCCCGTGTCT GCCTTCCACG TCAGCAACCG GAAGGACGGC
1551 TTCTGCATCA GTGGCAGCAT CACAGCCAAG TCCAGGTTTT ATGGGAACTC
1601 GCTGTCGGCG CTGCTGGACG GCAAAGCCAC GCTCACCTTC CTGAACCGAG
1651 CCGAGGATTA CACCCTTACC ATGCCCTACG CCCACTGCAA AGGAATCCTG
1701 TATGGCACGA TGACCCTGGA GCTGGGTGGG AAGGTCACCA TCGAGTGTGC
1751 GAAGAACAAC TTCCAGGCCC AGCTGGAATT CAAACTCAAG CCCTTCTTCG
1801 GGGGTAGCAC CAGCATCAAC CAGATCTCGG GAAAGATCAC GTCGGGAGAG
1851 GAAGTCCTGG CGAGCCTCAG TGGCCACTGG GACAGGGACG TGTTTATCAA
1901 GGAGGAAGGG AGCGGAAGCA GTGCGCTTTT CTGGACCCCG AGCGGGGAGG
1951 TCCGCAGACA GAGGCTGAGG CAGCACACGG TGCCGCTGGA GGAGCAGACG
2001 GAGCTGGAGT CCGAGAGGCT CTGGCAGCAC GTCACCAGGG CCATCAGCAA
2051 GGGCGACCAG CACAGGGCCA CACAGGAGAA GTTTGCACTG GAGGAGGCAC
2101 AGCGGCAGCG GGCCCGTGAG CGGCAGGAGA GCCTCATGCC CTGGAAGCCG
2151 CAGCTGTTCC ACCTGGACCC CATCACCCAG GAGTGGCACT ACCGATACGA
2201 GGACCACAGC CCCTGGGACC CCCTGAAGGA CATCGCCCAG TTTGAGCAAG
2251 ACGGGATCCT GCGGACCTTG CAGCAGGAGG CCGTGGCCCG CCAGACCACC
```

```
2301 TTCCTGGGCA GCCCAGGGCC CAGGCACGAG AGGTCTGGCC CAGACCAGCG
2351 GCTTCGCAAG GCCAGCGACC AGCCCTCCGG CCACAGCCAG GCCACGGAGA
2401 GCAGCGGATC CACGCCTGAG TCCTGCCCAG AGCTCTCAGA CGAGGAGCAG
2451 GATGGTGACT TTGTCCCTGG CGGTGAGAGC CCATGCCCTC GGTGCAGGAA
2501 GGAGGCGCGG CGGCTGCAGG CCCTGCACGA GGCCATCCTC TCCATCCGAG
2551 AGGCCCAGCA GGAGCTGCAC AGGCACCTCT CGGCCATGCT GAGCTCCACG
2601 GCACGGGCAG CACAGGCACC GACCCCAGGC CTCCTGCAGA GCCCCCGATC
2651 CTGGTTCCTG CTCTGCGTGT TCCTGGCGTG TCAGCTGTTC ATTAACCACA
2701 TCCTCAAATA GGAGCCCTGG GGGCAGAGCT CCTGGCCAGT CCCGAGCCCT
2751 CCCTCCCAGG CACCCAGCAC TTTAAGCCTG CTCCATGGAG GCAGAGAGGC
2801 CCGGCAAGCA CAGCCACTGT GACGGGGAGT CCAGGCGCAG GAGGGACCCG
2851 GGGCCACAAG GCGCTGCGGG CCCAGGTGTG CTGGGCCCCT CTCAGGGGCA
2901 CTGGCCTCTC TGCAGGGCCT TCCGCCCAGC GCTGGCCTTA ATGCTAAAGC
2951 CAAATGCAGC TTCTGCTGTG CGACGCACTC CTGGCCATCT TGCCGTGTCA
3001 CCCCCTGTCC GGCCTCCACT TGCCATGGGG GATGGATGGA TTTAGGGTGG
3051 GAGGGCCTGT GGGGGCCCTG GACAGTCACA CCCCAGCAGC AGTGAGTGGG
3101 CAGGTTTGGA GGAGCAGCCA GGGAGCCCCG AGTGGCCCAG GAGTCCCCCC
3151 ACACACAGAT GCATAGGCCT GCCTTCCGGA GACCCTGTCC ACATTGCCGG
3201 GACCACCCTG GTGGGGCCAC TGGTGGGTGC CAGGGACAGG TTAGGGCCAC
3251 TCTGGGGAAG GCATTTTGGT TTTTTATTCC ACGCTCTGCT GTTTGGATGG
3301 GAGCCCCACA GAGGCAGGTC CTGGAACCAC CCCACCCCCA CACCTGGACG
3351 CTCGCTCTGG TGGGGGCACA CGCAGGTGGA GGTGGTTGTG GGTGCAGGTG
3401 TGTGCAGGGG TGTGGGGGGC GCAGGGGTGT GGCTTAGCTG GCCCCGCACC
3451 CAGGCCGGGG AGGCTCAAGT TCGCCACTTT ACTCAGACCG ATGCACAGTC
3501 TTCCCATTTT ACACTTTTTT AATAAACATA ATTGCAATAT TTTAGGTGGG
3551 CTGCGAGCTG CAGTCAGCCT TCACGTCTGG CCTCAGTCCC CGTGTCAGTG
3601 CCGCTCTGCG TGTGCGTGTG CGCGTGTGTG AGCCTCTACA CATATATATA
3651 TGTACAGAGC CTTAAACCAC ATCGTGGCGG TGCCGTCTGA GCTGTAGCGG
3701 GTGGCTTTGT TTCCAGTTTT TGTACCCGTG TCCTTGTCTC CCCTCCTCCC
3751 CCATCTGGGG ATGTGTCTGT GTTCCACACC TTGAAATAAA CAGACACATA
3801 CGTGTTCTCT TAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

98315477:
The pleckstrin homology domain of oxysterol-binding protein recognises a determinant specific to Golgi membranes.

98146266:
A Drosophila homologue of oxysterol binding protein (OSBP)--implications for the role of OSBP.

98146266:
A Drosophila homologue of oxysterol binding protein (OSBP)--implications for the role of OSBP.

Peptide information for frame 3

ORF from 72 bp to 2708 bp; peptide length: 879
Category: strong similarity to known protein

```
  1 MKEEAFLRRR FSLCPPSSTP QKVDPRKLTR NLLLSGDNEL YPLSPGKDME
 51 PNGPSLPRDE GPPTPSSATK VPPAEYRLCN GSDKECVSPT ARVTKKETLK
101 AQKENYRQEK KRATRQLLSA LTDPSVVIMA DSLKIRGTLK SWTKLWCVLK
151 PGVLLIYKTP KVGQWVGTVL LHCCELIERP SKKDGFCFKL FHPLDQSVWA
201 VKGPKGESVG SITQPLPSSY LIFRAASESD GRCWLDALEL ALRCSSLLRL
251 GTCKPGRDGE PGTSPDASPS SLCGLPASAT VHPDQDLFPL NGSSLENDAF
301 SDKSERENPE ESDTETQDHS RKTESGSDQS ETPGAPVRRG TTYVEQVQEE
351 LGELGEASQV ETVSEENKSL MWTLLKQLRP GMDLSRVVLP TFVLEPRSFL
401 NKLSDYYYHA DLLSRAAVEE DAYSRMKLVL RWYLSGFYKK PKGIKKPYNP
451 ILGETFRCCW FHPQTDSRTF YIAEQVSHHP PVSAFHVSNR KDGFCISGSI
501 TAKSRFYGNS LSALLDGKAT LTFLNRAEDY TLTMPYAHCK GILYGTMTLE
551 LGGKVTIECA KNNFQAQLEF KLKPFFGGST SINQISGKIT SGEEVLASLS
601 GHWDRDVFIK EEGSGSSALF WTPSGEVRRQ RLRQHTVPLE EQTELESERL
```

```
651 WQHVTRAISK GDQHRATQEK FALEEAQRQR ARERQESLMP WKPQLFHLDP
701 ITQEWHYRYE DHSPWDPLKD IAQFEQDGIL RTLQQEAVAR QTTFLGSPGP
751 RHERSGPDQR LRKASDQPSG HSQATESSGS TPESCPELSD EEQDGDFVPG
801 GESPCPRCRK EARRLQALHE AILSIREAQQ ELHRHLSAML SSTARAAQAP
851 TPGLLQSPRS WFLLCVFLAC QLFINHILK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphute1_19h17, frame 3

TREMBL:CEZK1086_2 gene: "ZK1086.1"; Caenorhabditis elegans cosmid ZK1086, N = 1, Score = 1495, P = 2.7e-153

PIR:S25324 hypothetical protein YKR003w - yeast (Saccharomyces cerevisiae), N = 2, Score = 574, P = 8.5e-57

TREMBL:CEAF195_7 gene: "C32F10.1"; Caenorhabditis elegans cosmid C32F10., N = 1, Score = 588, P = 8.6e-57

PIR:S46796 hypothetical protein YKR003w homolog YHR001w - yeast (Saccharomyces cerevisiae), N = 1, Score = 585, P = 1.9e-56

TREMBL:NCOSBP_1 gene: "osbP"; product: "oxysterol-binding protein"; N.crassa mRNA for putative oxysterol-binding protein, N = 1, Score = 571, P = 7e-55

TREMBL:AB017026_1 product: "oxysterol-binding protein"; Mus musculus mRNA for oxysterol-binding protein, complete cds., N = 2, Score = 328, P = 3e-35

```
>TREMBL:CEZK1086_2 gene: "ZK1086.1"; Caenorhabditis elegans cosmid ZK1086
          Length = 751

  HSPs:

 Score = 1495 (224.3 bits), Expect = 2.7e-153, P = 2.7e-153
 Identities = 327/663 (49%), Positives = 430/663 (64%)

Query:   129 MADSLKIRGTLKSWTKLWCVLKPGVLLIYKTPKV--GQWVGTVLLHCCELIERPSKKDGF 186
             MAD+LKIRG LK W + +CVLKPG+L++YK  K   G WVGTVLL+ CELIERPSKKDGF
Sbjct:     1 MADTLKIRGALKRWNRYYCVLKPGLLILYKHKKADRGDWVGTVLLNHCELIERPSKKDGF 60

Query:   187 CFKLFHPLDQSVWAVKGPKGESVGSIT-QPLPSSYLIFRAASESDGRCWLDALELALRCS 245
             CFKLFHP+D S+W  +GP G+S GS T  PL +S+LI RA S+  GRCW+DALEL+ +C+
Sbjct:    61 CFKLFHPMDMSIWGNRGPLGQSFGSFTLNPLNTSFLICRAPSDQAGRCWMDALELSFKCT 120

Query:   246 SLLRLGTCKPGRDGEPGTSPDASPSSLCGLPASATVHPDQDLFPLNGSSLENDAFSDK-S 304
             LL+  T     D + G   D+S +   G    + + D D      G     A S+  +
Sbjct:   121 GLLKK-TMNE-LDDKNG---DSSMND--GQRDESRMSRDSD-----GDDTRELAVSETDA 168

Query:   305 ERENPEESDTETQDHSRKTESGSDQSETPGAPVRRGTT---YVEQVQEELGELGEASQVE 361
             E+    E  D + +DH    E G   SET   +R   T   ++   +E  G  G  S  E
Sbjct:   169 EKHFQEIDDVQDEDH----EDGK-MSETSDT-IREAFTESAWIPSPKEVFGPDG--SLTE 220

Query:   362 TVSEENKSLMWTLLKQLRPGMDLSRVVLPTFVLEPRSFLNKLSDYYYHADLLSRAAVEED 421
             V EENKSL+WTLLKQ+RPGMDLS+VVLPTF+LEPRSFL KL+DYYYHADL+S A  E D
Sbjct:   221 EVGEENKSLIWTLLKQIRPGMDLSKVVLPTFILEPRSFLEKLADYYYHADLISEAVAEPD 280

Query:   422 AYSRMKLVLRWYLSGFYKKPKGIKKPYNPILGETFRCCWFHPQTDSRTFYIAEQVSHHPP 481
              + R+  V +++LSGFYKKPKG+KKPYNPILGETFRC W HP    S TFY+AEQVSHHPP
Sbjct:   281 PFQRIVKVTKFFLSGFYKKPKGLKKPYNPILGETFRCKWEHPD-GSTTFYMAEQVSHHPP 339

Query:   482 VSAFHVSNRKDGFCISGSITAKSRFYGNSLSALLDGKATLTFLNRAEDYTLTMPYAHCKG 541
             VS+  ++NRK GF ISG+I AKS++YGNSLSA+L GK  LT LN  E Y + +PYA+CKG
Sbjct:   340 VSSLFITNRKAGFNISGTILAKSKYYGNSLSAILAGKLRLTLLNLGETYIVNLPYANCKG 399

Query:   542 ILYGTMTLELGGKVTIECAKNNFQAQLEFKLKPFFGGSTSINQISGKITSGEEVLASLSG 601
             I+ GTMT+ELGG+V IEC K  ++  L+FKLKP  GG+   NQI G I  G + LAS+ G
Sbjct:   400 IMIGTMTMELGGEVNIECEKTGYRTTLDFKLKPMLGGA--YNQIEGSIKYGSDRLASIEG 457

Query:   602 HWDRDVFIKEEGSGSSALFWTPSGEVRRQRLRQHTVPLEEQTELESERLWQHVTRAISKG 661
              WD  + IK  G        W P+ EV + RL ++ + ++EQ E ES +LW+HVT AIS
Sbjct:   458 AWDGVIRIK--GPDGKKELWNPTPEVIKTRLPRYEINMDEQGEWESAKLWRHVTEAISNE 515

Query:   662 DQHRATQEKFALEEAQRQRARERQESLMPWKPQLFHLDPITQEWHYRYEDHSPWDPLKDI 721
             DQ++AT+EK ALE  QR RA+    S +P + + F       ++ Y + D+ PWD   DI
Sbjct:   516 DQYKATEEKTALENDQRARAK----SGIPHETKFFKKQH-GDDYVYIHADYRPWDNNNDI 570
```

```
Query:   722 AQFEQDGILRTLQQEAVAR--QTTFLGSPGPRHERSGPDQRLRKASDQPSGHSQATESSG 779
             Q E + +++T+ + +  +   + LGS     E S  D+ +     +P    +    +
Sbjct:   571 QQIENNYVVKTISRHSKRKTGNSEQLGSDNTS-EASESDEEVI----EPKIKKKEIVPAK 625

Query:   780 STPESCPELSDE 791
             S P + PE++DE
Sbjct:   626 SKPIT-PEVADE 636
```

Pedant information for DKFZphute1_19h17, frame 3

Report for DKFZphute1_19h17.3

```
[LENGTH]    879
[MW]        98616.79
[pI]        7.29
[HOMOL]     TREMBL:CEZK1086_2 gene: "ZK1086.1"; Caenorhabditis elegans cosmid ZK1086 1e-157

[FUNCAT]    01.06.16 lipid and fatty-acid binding      [S. cerevisiae, YHR001w] 3e-55
[FUNCAT]    01.06.01 lipid, fatty-acid and sterol biosynthesis   [S. cerevisiae, YHR001w]
3e-55
[FUNCAT]    30.03 organization of cytoplasm     [S. cerevisiae, YPL145c] 3e-23
[FUNCAT]    08.07 vesicular transport (golgi network, etc.)     [S. cerevisiae, YPL145c]
3e-23
[FUNCAT]    04.05.01.07 chromatin modification    [S. cerevisiae, YAR044w] 5e-20
[BLOCKS]    BL00168F
[BLOCKS]    BL01013D Oxysterol-binding protein family proteins
[BLOCKS]    BL01013C Oxysterol-binding protein family proteins
[BLOCKS]    BL01013B Oxysterol-binding protein family proteins
[BLOCKS]    BL01013A Oxysterol-binding protein family proteins
[PIRKW]     transmembrane protein 1e-19
[SUPFAM]    pleckstrin repeat homology 8e-18
[SUPFAM]    ankyrin repeat homology 1e-19
[SUPFAM]    unassigned ankyrin repeat proteins 1e-19
[PROSITE]   MYRISTYL      12
[PROSITE]   CAMP_PHOSPHO_SITE    6
[PROSITE]   OSBP    1
[PROSITE]   CK2_PHOSPHO_SITE     21
[PROSITE]   PROKAR_LIPOPROTEIN   1
[PROSITE]   TYR_PHOSPHO_SITE     2
[PROSITE]   PKC_PHOSPHO_SITE     20
[PROSITE]   ASN_GLYCOSYLATION    3
[PFAM]      PH (pleckstrin  homology) domain
[KW]        TRANSMEMBRANE  1
[KW]        LOW_COMPLEXITY     2.96 %
[KW]        COILED_COIL        3.53 %
```

```
SEQ    MKEEAFLRRRFSLCPPSSTPQKVDPRKLTRNLLLSGDNELYPLSPGKDMEPNGPSLPRDE
SEG    ............................................................
PRD    ccchhhhhhhhccccccccccccccccccccccccccccccccccccccccccccccccc
COILS  ............................................................
MEM    ............................................................

SEQ    GPPTPSSATKVPPAEYRLCNGSDKECVSPTARVTKKETLKAQKENYRQEKKRATRQLLSA
SEG    ............................................................
PRD    cccccccccccccceeeeccccccceeeeccccchhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  ...............................ccccccccccccccccccccccccccccc
MEM    ............................................................

SEQ    LTDPSVVIMADSLKIRGTLKSWTKLWCVLKPGVLLIYKTPKVGQWVGTVLLHCCELIERP
SEG    ............................................................
PRD    hccccceeeeccccccccccccccceeeeeecccceeeeeccccccccceeeeecccccccccc
COILS  CCC.........................................................
MEM    ............................................................

SEQ    SKKDGFCFKLFHPLDQSVWAVKGPKGESVGSITQPLPSSYLIFRAASESDGRCWLDALEL
SEG    ............................................................
PRD    cccccceeeeeccccccceeeeeccccccceeeccccccccceeeeeeehhhhhhhhhhhhhh
COILS  ............................................................
MEM    ............................................................

SEQ    ALRCSSLLRLGTCKPGRDGEPGTSPDASPSSLCGLPASATVHPDQDLFPLNGSSLENDAF
SEG    ............................................................
PRD    hhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
COILS  ............................................................
MEM    ............................................................

SEQ    SDKSERENPEESDTETQDHSRKTESGSDQSETPGAPVRRGTTYVEQVQEELGELGEASQV
```

```
SEG    ...........................................xxxxxxxxxxxxx....
PRD    cccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhcccccc
COILS  ............................................................
MEM    ............................................................

SEQ    ETVSEENKSLMWTLLKQLRPGMDLSRVVLPTFVLEPRSFLNKLSDYYYHADLLSRAAVEE
SEG    ............................................................
PRD    ccccccchhhhhhhhhcccccceeeccccceeeeccccchhhhhhhhhcccccccccccc
COILS  ............................................................
MEM    ............................................................

SEQ    DAYSRMKLVLRWYLSGFYKKPKGIKKPYNPILGETFRCCWFHPQTDSRTFYIAEQVSHHP
SEG    ............................................................
PRD    chhhhhhhhhhhhhhhcccccccccccccccccceeeeeeccccccccceeeeecccccc
COILS  ............................................................
MEM    ............................................................

SEQ    PVSAFHVSNRKDGFCISGSITAKSRFYGNSLSALLDGKATLTFLNRAEDYTLTMPYAHCK
SEG    ............................................................
PRD    cceeeeeccccccccccccccccccccccccccccccceeeeeecccceeeeccccceee
COILS  ............................................................
MEM    ............................................................

SEQ    GILYGTMTLELGGKVTIECAKNNFQAQLEFKLKPFFGGSTSINQISGKITSGEEVLASLS
SEG    ............................................................
PRD    eeeeccccccccceeeeecccccccceeeeccccccccccccceeeeeccccccceeeeec
COILS  ............................................................
MEM    ............................................................

SEQ    GHWDRDVFIKEEGSGSSALFWTPSGEVRRQRLRQHTVPLEEQTELESERLWQHVTRAISK
SEG    ............................................................
PRD    cccccceeeeecccccceeeeeccccccccccccccccccccchhhhhhhhhhhhhhhhhh
COILS  ............................................................
MEM    ............................................................

SEQ    GDQHRATQEKFALEEAQRQRARERQESLMPWKPQLFHLDPITQEWHYRYEDHSPWDPLKD
SEG    .............xxxxxxxxxxxxx..................................
PRD    cchhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccceeeecccccccchh
COILS  ............................................................
MEM    ............................................................

SEQ    IAQFEQDGILRTLQQEAVARQTTFLGSPGPRHERSGPDQRLRKASDQPSGHSQATESSGS
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhhhhcccccccccccccchhhhcccccccccccccccccccc
COILS  ............................................................
MEM    ............................................................

SEQ    TPESCPELSDEEQDGDFVPGGESPCPRCRKEARRLQALHEAILSIREAQQELHRHLSAML
SEG    ............................................................
PRD    cccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  ............................................................
MEM    ............................................................

SEQ    SSTARAAQAPTPGLLQSPRSWFLLCVFLACQLFINHILK
SEG    .......................................
PRD    hhhhhhccccccccccccceeeeehhhhhhhhhhhhccc
COILS  .......................................
MEM    .....................MMMMMMMMMMMMMMMMM.
```

Prosite for DKFZphutel_19h17.3

PS00001	80->84	ASN_GLYCOSYLATION	PDOC00001
PS00001	291->295	ASN_GLYCOSYLATION	PDOC00001
PS00001	367->371	ASN_GLYCOSYLATION	PDOC00001
PS00004	9->13	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	26->30	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	95->99	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	111->115	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	338->342	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	762->766	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	90->93	PKC_PHOSPHO_SITE	PDOC00005
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	98->101	PKC_PHOSPHO_SITE	PDOC00005
PS00005	132->135	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	159->162	PKC_PHOSPHO_SITE	PDOC00005
PS00005	181->184	PKC_PHOSPHO_SITE	PDOC00005
PS00005	252->255	PKC_PHOSPHO_SITE	PDOC00005

PS00005	301->304	PKC_PHOSPHO_SITE	PDOC00005
PS00005	304->307	PKC_PHOSPHO_SITE	PDOC00005
PS00005	320->323	PKC_PHOSPHO_SITE	PDOC00005
PS00005	455->458	PKC_PHOSPHO_SITE	PDOC00005
PS00005	488->491	PKC_PHOSPHO_SITE	PDOC00005
PS00005	501->504	PKC_PHOSPHO_SITE	PDOC00005
PS00005	586->589	PKC_PHOSPHO_SITE	PDOC00005
PS00005	647->650	PKC_PHOSPHO_SITE	PDOC00005
PS00005	824->827	PKC_PHOSPHO_SITE	PDOC00005
PS00005	843->846	PKC_PHOSPHO_SITE	PDOC00005
PS00005	857->860	PKC_PHOSPHO_SITE	PDOC00005
PS00006	82->86	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	227->231	CK2_PHOSPHO_SITE	PDOC00006
PS00006	263->267	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00006	312->316	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00006	362->366	CK2_PHOSPHO_SITE	PDOC00006
PS00006	590->594	CK2_PHOSPHO_SITE	PDOC00006
PS00006	643->647	CK2_PHOSPHO_SITE	PDOC00006
PS00006	659->663	CK2_PHOSPHO_SITE	PDOC00006
PS00006	713->717	CK2_PHOSPHO_SITE	PDOC00006
PS00006	755->759	CK2_PHOSPHO_SITE	PDOC00006
PS00006	780->784	CK2_PHOSPHO_SITE	PDOC00006
PS00006	784->788	CK2_PHOSPHO_SITE	PDOC00006
PS00006	789->793	CK2_PHOSPHO_SITE	PDOC00006
PS00006	824->828	CK2_PHOSPHO_SITE	PDOC00006
PS00007	402->409	TYR_PHOSPHO_SITE	PDOC00007
PS00007	415->424	TYR_PHOSPHO_SITE	PDOC00007
PS00008	137->143	MYRISTYL	PDOC00008
PS00008	163->169	MYRISTYL	PDOC00008
PS00008	274->280	MYRISTYL	PDOC00008
PS00008	326->332	MYRISTYL	PDOC00008
PS00008	381->387	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	508->514	MYRISTYL	PDOC00008
PS00008	541->547	MYRISTYL	PDOC00008
PS00008	552->558	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	613->619	MYRISTYL	PDOC00008
PS00008	728->734	MYRISTYL	PDOC00008
PS00013	860->871	PROKAR_LIPOPROTEIN	PDOC00013
PS01013	474->485	OSBP	PDOC00774

Pfam for DKFZphutel_19h17.3

```
HMM_NAME       PH (pleckstrin  homology) domain

HMM                *dvIREGWMyKWgswrkstgnWqrRWFvLrndpnrLiYYkddkdekPrYM
                    +VI+  +++++G  +     W + W+VL++  ++L+ YK + +  + ++
Query          126  VVIMADSLKIRGTLKS----WTKLWCVLKP--GVLLIYKTP-KVGQWVG   167

HMM                lIdldcWrMidVEidWmmdndHCFiIWtrq.................
                      L+C+ +I+ ++ ++    +CF+++ +
Query          168 TVLLHCCELIERPSKKD---GFCFKLFHPLDQSVWAVKGPKGESVGSITQ   214

HMM                ....rtYYFQAeNeEEMmeWMsaIrRaIw*
                       + ++F+A++E++ + W++A++ A++
Query          215 PLPSSYLIFRAASESDGRCWLDALELALR   243
```

DKFZphutel_19j11
----

group: uterus derived

DKFZphutel_19j11 encodes a novel 708 amino acid protein with C-terminal similarity to several known proteins, such as human KIAA0231 or murine ras binding protein Sur8.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

Strong similarity to KIAA0231, similarity to ras binding protein Sur8

EST AA854189 extendes the sequence (294 Bp), with this sequence complete cDNA,

Sequenced by AGOWA

Locus: unknown

Insert length: 2343 bp
Poly A stretch at pos. 2323, polyadenylation signal at pos. 2295

```
   1 GCTCCTGCTA ACCCCATCAC TGTGGAAATG AAAGGCCTGA AGACAGATTT
  51 GGACCTTCAG CAGTACAGCT TTATAAATCA GATGTGTTAT GAGCGAGCCC
 101 TCCACTGGTA TGCCAAGTAT TTCCCTTACC TTGTCCTCAT CCATACCCTG
 151 GTCTTTATGC TCTGCAGTAA CTTTTGGTTC AAATTCCCTG GTTCCAGCTC
 201 CAAAATAGAA CATTTCATCT CCATTCTGGG GAAGTGTTTT GACTCTCCTT
 251 GGACCACACG GGCTTTATCT GAAGTGTCTG GGGAGGACTC AGAAGAAAAG
 301 GACAACAGGA AGAACAACAT GAACAGGTCC AACACCATCC AATCTGGTCC
 351 AGAAGGCAGC CTGGTCAACT CTCAGTCTTT AAAGTCCATT CCTGAGAAGT
 401 TTGTAGTTGA TAAATCCACT GCAGGGGCTC TGGATAAAAA GGAAGGTGAG
 451 CAGGCTAAGG CCTTATTTGA GAAGGTGAAG AAGTTCAGGC TGCATGTGGA
 501 AGAAGGTGAT ATTCTATATG CCATGTATGT TCGCCAGACT GTACTTAAAG
 551 TTATCAAATT CCTAATCATC ATTGCATATA ATAGTGCTCT GGTTTCCAAG
 601 GTCCAGTTTA CAGTGGACTG TAATGTGGAC ATTCAGGACA TGACTGGATA
 651 TAAAAACTTT TCTTGCAATC ATACCATGGC ACACTTGTTC TCAAAACTGT
 701 CCTTTTGCTA TCTGTGCTTT GTTAGTATCT ATGGATTGAC GTGCCTTTAT
 751 ACCTTATACT GGCTGTTCTA CCGTTCTCTA CGGGAATATT CCTTTGAGTA
 801 TGTCCGTCAG GAGACTGGAA TTGATGATAT TCCAGATGTG AAAAATGACT
 851 TTGCTTTTAT GCTTCATATG ATAGATCAGT ATGACCCTCT CTATTCCAAG
 901 AGATTTGCAG TGTTCCTGTC TGAAGTCAGT GAAAACAAAT TAAAGCAGCT
 951 GAACTTAAAT AACGAATGGA CTCCTGATAA ACTGAGGCAG AAGCTACAGA
1001 CAAATGCCCA TAATCGACTG GAATTGCCTC TTATCATGCT CTCTGGCCTT
1051 CCAGACACTG TTTTTGAAAT CACAGAGTTG CAATCTCTAA AACTTGAAAT
1101 CATTAAGAAC GTAATGATAC CAGCCACCAT TGCACAGCTA GACAATCTTC
1151 AAGAGCTCTC TCTGCACCAG TGTTCTGTCA AAATCCACAG TGCGGCGCTC
1201 TCTTTCCTGA AGGAAAACCT CAAGGTCTTG AGCGTCAAGT TTGATGACAT
1251 GAGGGAACTC CCCCCCTGGA TGTATGGGCT CCGAAATCTG GAAGAGCTGT
1301 ACCTAGTTGG CTCTCTAAGT CATGATATTT CCAGAAATGT CACCCTTGAG
1351 TCTCTGCGGG ATCTCAAAAG CCTTAAAATT CTCTCTATCA AAAGCAACGT
1401 TTCCAAAATC CCTCAGGCAG TGGTTGATGT TTCCAGCCAT CTCCAGAAGA
1451 TGTGCATACA TAATGATGGC ACCAAGCTGG TGATGCTCAA CAACTTAAAG
1501 AAGATGACCA ATCTGACAGA GCTGGAGCTG GTCCACTGTG ACCTGGAGCG
1551 TATTCCTCAT GCTGTGTTCA GCCTACTCAG CCTCCAGGAA TTGGACCTGA
1601 AGGAAAACAA TCTGAAATCT ATAGAAGAAA TCGTTAGCTT TCAGCACTTA
1651 AGAAAGTTGA CAGTGCTAAA ACTGTGGCAT AACAGCATCA CCTACATCCC
1701 AGAGCATATA AAGAAACTCA CCAGCCTGGA ACGCCTGTCC TTTAGTCACA
1751 ATAAAATAGA GGTGCTGCCT TCCCACCTCT TCCTATGCAA CAAGATCCGA
1801 TACTTGGACT TATCGTACAA TGACATTCGA TTTATCCCCC CTGAAATTGG
1851 AGTTCTACAA AGTTTACAGT ATTTTTCCAT CACATGTAAC AAAGTGGAAA
1901 GCCTTCCAGA TGAACTCTAC TTCTGCAAGA AACTTAAAAC TCTGAAGATT
1951 GGAAAAAACA GCCTATCTGT ACTTTCACCG AAAATTGGAA ATTTGCTATT
2001 TCTTTCCTAC TTAGATGTAA AAGGTAATCA CTTTGAAATC CTCCCTCCTG
2051 AACTGGGTGA CTGTCGGGCT CTGAAGCGAG CTGGTTTAGT TGTAGAAGAT
2101 GCTCTGTTTG AAACTCTGCC TTCTGACGTC CGGGAGCAAA TGAAAACAGA
2151 ATAACTTATT TTTCGTTAAA GTTTGACTGA AACACGCTTC TACCAAATAC
2201 AGTATAAATA ATTAGGTAGT CTTAATGCCT TTCCTATTTT TTTTTCCTTT
2251 TCACACAAAA TGTACACAAA GATCGCGTAA GGAGTATGTA TTTTTAATAA
2301 AAATTTAATT GTATTTTTTC AATATTAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

96421675:
Characterization of densin-180, a new brain-specific synaptic protein of the
O-sialoglycoprotein family.

98337190:
SUR-8, a conserved Ras-binding protein with leucine-rich repeats, positively regulates Ras-mediated signaling in C. elegans.

Peptide information for frame 1

ORF from 28 bp to 2151 bp; peptide length: 708
Category: similarity to known protein
Classification: Cell signaling/communication

```
  1 MKGLKTDLDL QQYSFINQMC YERALHWYAK YFPYLVLIHT LVFMLCSNFW
 51 FKFPGSSSKI EHFISILGKC FDSPWTTRAL SEVSGEDSEE KDNRKNNMNR
101 SNTIQSGPEG SLVNSQSLKS IPEKFVVDKS TAGALDKKEG EQAKALFEKV
151 KKFRLHVEEG DILYAMYVRQ TVLKVIKFLI IIAYNSALVS KVQFTVDCNV
201 DIQDMTGYKN FSCNHTMAHL FSKLSFCYLC FVSIYGLTCL YTLYWLFYRS
251 LREYSFEYVR QETGIDDIPD VKNDFAFMLH MIDQYDPLYS KRFAVFLSEV
301 SENKLKQLNL NNEWTPDKLR QKLQTNAHNR LELPLIMLSG LPDTVFEITE
351 LQSLKLEIIK NVMIPATIAQ LDNLQELSLH QCSVKIHSAA LSFLKENLKV
401 LSVKFDDMRE LPPWMYGLRN LEELYLVGSL SHDISRNVTL ESLRDLKSLK
451 ILSIKSNVSK IPQAVVDVSS HLQKMCIHND GTKLVMLNNL KKMTNLTELE
501 LVHCDLERIP HAVFSLLSLQ ELDLKENNLK SIEEIVSFQH LRKLTVLKLW
551 HNSITYIPEH IKKLTSLERL SFSHNKIEVL PSHLFLCNKI RYLDLSYNDI
601 RFIPPEIGVL QSLQYFSITC NKVESLPDEL YFCKKLKTLK IGKNSLSVLS
651 PKIGNLLFLS YLDVKGNHFE ILPPELGDCR ALKRAGLVVE DALFETLPSD
701 VREQMKTE
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphute1_19j11, frame 1

TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene, partial cds., N = 1, Score = 1408, P = 4.5e-144

TREMBL:AF054827_1 gene: "soc-2"; product: "leucine-rich repeat protein SOC-2"; Caenorhabditis elegans leucine-rich repeat protein SOC-2 (soc-2) mRNA, complete cds., N = 1, Score = 304, P = 5.7e-24

TREMBL:RNU66707_1 product: "densin-180"; Rattus norvegicus densin-180 mRNA, complete cds., N = 1, Score = 311, P = 7.4e-24

TREMBL:AF068921_1 product: "Ras-binding protein SUR-8"; Mus musculus Ras-binding protein SUR-8 mRNA, complete cds., N = 1, Score = 302, P = 1.1e-23

>TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene, partial cds.
Length = 476

HSPs:

Score = 1408 (211.3 bits), Expect = 4.5e-144, P = 4.5e-144
Identities = 265/471 (56%), Positives = 361/471 (76%)

```
Query:  237 LTCLYTLYWLFYRSLREYSFEYVRQETGIDDIPDVKNDFAFMLHMIDQYDPLYSKRFAVF 296
            LT  Y+L+W+    SL++YSFE +R+++    DIPDVKNDFAF+LH+ DQYDPLYSKRF++F
Sbjct:    1 LTSSYSLWWMLRSSLKQYSFEALREKSNYSDIPDVKNDFAFILHLADQYDPLYSKRFSIF 60

Query:  297 LSEVSENKLKQLNLNNEWTPDKLRQKLQTNAHNRLELPLIMLSGLPDTVFEITELQSLKL 356
            LSEVSENKLKQ+NLNNEWT +KL+ KL  NA +++EL L ML+GLPD VFE+TE++ L L
Sbjct:   61 LSEVSENKLKQINLNNEWTVEKLKSKLVKNAQDKIELHLFMLNGLPDNVFELTEMEVLSL 120
```

```
Query:   357 EIIKNVMIPATIAQLDNLQELSLHQCSVKIHSAALSFLKENLKVLSVKFDDMRELPPWMY 416
             E+I  V +P+ ++QL NL+EL ++  S+ +   AL+FL+ENLK+L +KF +M ++P W++
Sbjct:   121 ELIPEVKLPSAVSQLVNLKELRVYHSSLVVDHPALAFLEENLKILRLKFTEMGKIPRWVF 180

Query:   417 GLRNLEELYLVGSLSHDISRNVTLESLRDLKSLKILSIKSNVSKIPQAVVDVSSHLQKMC 476
             L+NL+ELYL G +  +     + LE  +DLK+L+ L +KS++S+IPQ V D+    LQK+
Sbjct:   181 HLKNLKELYLSGCVLPEQLSTMQLEGFQDLKNLRTLYLKSSLSRIPQVVTDLLPSLQKLS 240

Query:   477 IHNDGTKLVMLNNLKKMTNLTELELVHCDLERIPHAVFSLLSLQELDLKENNLKSIEEIV 536
             + N+G+KLV+LNNLKKM NL  LEL+ CDLERIPH++FSL +L ELDL+ENNLK++EEI+
Sbjct:   241 LDNEGSKLVVLNNLKKMVNLKSLELISCDLERIPHSIFSLNNLHELDLRENNLKTVEEII 300

Query:   537 SFQHLRKLTVLKLWHNSITYIPEHIKKLTSLERLSFSHNKIEVLPSHLFLCNKIRYLDLS 596
             SFQHL+ L+ LKLWHN+I YIP  I  L++LE+LS  HN IE LP  LFLC K+ YLDLS
Sbjct:   301 SFQHLQNLSCLKLWHNNIAYIPAQIGALSNLEQLSLDHNNIENLPLQLFLCTKLHYLDLS 360

Query:   597 YNDIRFIPPEIGVLQSLQYFSITCNKVESLPDELYFCKKLKTLKIGKNSLSVLSPKIGNL 656
             YN + FIP EI  L +LQYF++T N +E LPD L+ CKKL+ L +GKNSL  LSP +G L
Sbjct:   361 YNHLTFIPEEIQYLSNLQYFAVTNNNIEMLPDGLFQCKKLQCLLLGKNSLMNLSPHVGEL 420

Query:   657 LFLSYLDVKGNHFEILPPELGDCRALKRAGLVVEDALFETLPSDVREQMKT 707
             L++L++ GN+ E LPPEL  C++LKR  L+VE+ L  TLP  V E+++T
Sbjct:   421 SNLTHLELIGNYLETLPPELEGCQSLKRNCLIVEENLLNTLPLPVTERLQT 471
```

Pedant information for DKFZphutel_19j11, frame 1

Report for DKFZphutel_19j11.1

```
[LENGTH]        708
[MW]            81812.82
[pI]            7.55
[HOMOL]         TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene, partial cds.
1e-149
[FUNCAT]        30.02 organization of plasma membrane    [S. cerevisiae, YJL005w] 3e-17
[FUNCAT]        03.22 cell cycle control and mitosis [S. cerevisiae, YJL005w] 3e-17
[FUNCAT]        10.04.03 second messenger formation  [S. cerevisiae, YJL005w] 3e-17
[FUNCAT]        01.03.10 metabolism of cyclic and unusual nucleotides     [S. cerevisiae,
YJL005w] 3e-17
[FUNCAT]        03.10 sporulation and germination    [S. cerevisiae, YJL005w] 3e-17
[FUNCAT]        30.10 nuclear organization   [S. cerevisiae, YKL193c] 3e-09
[FUNCAT]        06.07 protein modification (glycolsylation, acylation, myristylation,
palmitylation, farnesylation and processing)   [S. cerevisiae, YKL193c] 3e-09
[FUNCAT]        04.05.01.04 transcriptional control  [S. cerevisiae, YAL021c] 9e-08
[FUNCAT]        01.05.04 regulation of carbohydrate utilization     [S. cerevisiae, YAL021c]
9e-08
[FUNCAT]        01.01.04 regulation of amino-acid metabolism       [S. cerevisiae, YAL021c]
9e-08
[FUNCAT]        99 unclassified proteins     [S. cerevisiae, YOR353c] 3e-07
[BLOCKS]        BL00868F
[BLOCKS]        BL00985B Spermadhesins family proteins
[EC]            3.4.17.3 Lysine carboxypeptidase 1e-08
[EC]            4.6.1.1 Adenylate cyclase 3e-18
[PIRKW]         blocked amino end 1e-10
[PIRKW]         phosphotransferase 1e-09
[PIRKW]         nucleus 6e-08
[PIRKW]         duplication 3e-18
[PIRKW]         platelet 1e-10
[PIRKW]         tandem repeat 7e-16
[PIRKW]         keratan sulfate 7e-07
[PIRKW]         metallo-carboxypeptidase 1e-08
[PIRKW]         transmembrane protein 1e-10
[PIRKW]         serine/threonine-specific protein kinase 1e-09
[PIRKW]         autophosphorylation 1e-09
[PIRKW]         cartilage 7e-07
[PIRKW]         connective tissue 7e-07
[PIRKW]         magnesium 1e-09
[PIRKW]         cAMP biosynthesis 3e-18
[PIRKW]         ATP 1e-09
[PIRKW]         receptor 1e-09
[PIRKW]         leucine zipper 3e-13
[PIRKW]         glycoprotein 5e-12
[PIRKW]         extracellular matrix 7e-07
[PIRKW]         chondroitin sulfate proteoglycan 7e-07
[PIRKW]         cell adhesion 1e-08
[PIRKW]         hydrolase 1e-08
[PIRKW]         sulfoprotein 7e-07
[PIRKW]         membrane protein 1e-08
[PIRKW]         phosphorus-oxygen lyase 3e-18
```

```
[PIRKW]         collagen binding 7e-07
[SUPFAM]        leucine-rich alpha-2-glycoprotein repeat homology 3e-21
[SUPFAM]        chaoptin 1e-08
[SUPFAM]        gelsolin repeat homology 3e-21
[SUPFAM]        protein kinase homology 1e-09
[SUPFAM]        protein kinase Xa21 1e-09
[SUPFAM]        fibromodulin 4e-12
[SUPFAM]        yeast adenylate cyclase catalytic domain homology 3e-18
[SUPFAM]        yeast adenylate cyclase 3e-18
[KW]            TRANSMEMBRANE  3
[KW]            LOW_COMPLEXITY     1.41 %

SEQ   MKGLKTDLDLQQYSFINQMCYERALHWYAKYFPYLVLIHTLVFMLCSNFWFKFPGSSSKI
SEG   ............................................................
PRD   ccccchhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhccceeeeccccccee
MEM   ..................................MMMMMMMMMMMMMMMMM.........

SEQ   EHFISILGKCFDSPWTTRALSEVSGEDSEEKDNRKNNMNRSNTIQSGPEGSLVNSQSLKS
SEG   ............................................................
PRD   eeeeeeecccccccccceeeeeccccccccccccccccccccccccccccceeeeccccc
MEM   ............................................................

SEQ   IPEKFVVDKSTAGALDKKEGEQAKALFEKVKKFRLHVEEGDILYAMYVRQTVLKVIKFLI
SEG   ............................................................
PRD   cccceeeccccccchhhhhhhhhhhhhhhhhhhhhhhhcccceeeehhhhhhhhhhhhhh
MEM   ...................................................MMMMMMMMM

SEQ   IIAYNSALVSKVQFTVDCNVDIQDMTGYKNFSCNHTMAHLFSKLSFCYLCFVSIYGLTCL
SEG   ............................................................
PRD   hhhhcchhhheeeeecccccccccccccccccccchhhhhhhheeeeeeeeeeeccceee
MEM   MMMMMMMM...................................MMMMMMMMMMMMMMMMM

SEQ   YTLYWLFYRSLREYSFEYVRQETGIDDIPDVKNDFAFMLHMIDQYDPLYSKRFAVFLSEV
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhcccccccccccchhhhhhhhhccchhhhhhhhhhhhhhh
MEM   ............................................................

SEQ   SENKLKQLNLNNEWTPDKLRQKLQTNAHNRLELPLIMLSGLPDTVFEITELQSLKLEIIK
SEG   ..xxxxxxxxxx................................................
PRD   hhhhhhhhccccccccchhhhhhhhhhhhhhhhhhccccccchhhhhhhhhhhhhhhhhh
MEM   ............................................................

SEQ   NVMIPATIAQLDNLQELSLHQCSVKIHSAALSFLKENLKVLSVKFDDMRELPPWMYGLRN
SEG   ............................................................
PRD   hcccccchhhhhhhhhhhhhccccccccccccchhhhhhhhhcccccccccccccchhhh
MEM   ............................................................

SEQ   LEELYLVGSLSHDISRNVTLESLRDLKSLKILSIKSNVSKIPQAVVDVSSHLQKMCIHND
SEG   ............................................................
PRD   hhhhhhccccccccccccccchhhhhhhhhhhcccccccccccchhhhhhhhhhhhcccc
MEM   ............................................................

SEQ   GTKLVMLNNLKKMTNLTELELVHCDLERIPHAVFSLLSLQELDLKENNLKSIEEIVSFQH
SEG   ............................................................
PRD   ceeeeccccccchhhhhhhhcccccccccccchhhhhhhhhhhcccccccccccccccch
MEM   ............................................................

SEQ   LRKLTVLKLWHNSITYIPEHIKKLTSLERLSFSHNKIEVLPSHLFLCNKIRYLDLSYNDI
SEG   ............................................................
PRD   hhhhhhcccccceeecccccchhhhheeeccccceeecccccchhhhhhhhhhhhccccc
MEM   ............................................................

SEQ   RFIPPEIGVLQSLQYFSITCNKVESLPDELYFCKKLKTLKIGKNSLSVLSPKIGNLLFLS
SEG   ............................................................
PRD   cccccccchhhhhhhhhhhcccccccccccccchhhhcccccccccceecccccccchhhh
MEM   ............................................................

SEQ   YLDVKGNHFEILPPELGDCRALKRAGLVVEDALFETLPSDVREQMKTE
SEG   ................................................
PRD   hhhcccccccccccccchhhhhhhheeeecccccccccccccccccccc
MEM   ................................................
```

(No Prosite data available for DKFZphutel_19j11.1)

(No Pfam data available for DKFZphutel_19j11.1)

DKFZphute1_1i2

group: transcription factor

DKFZphute1_1i2 encodes a novel 594 amino acid protein similar to signal transducing proteins.

The protein contains 2 WD-40 repeats, which is typical for the beta-transducin subunit of G-proteins. In addition, the protein contains a C3HC4 zinc finger and a leucine zipper. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. Due to the zinc finger the novel protein seems to be a new molecule involved in signal transduction and transcription.

The new protein can find application in modulating/blocking gene expression of genes controlled by this molecule.

similarity to Dictostelium myosin heavy chain kinase

complete cDNA, complete cds, EST hits
[PFAM] Zinc finger, C3HC4 type (RING finger)
[PFAM] WD domain, G-beta repeats
[SCOP] d1tbgc_ 2.46.3.1.1 beta1-subunit of the signal-transducing G protei 3e-07

Sequenced by BMFZ

Locus: /map="16p13.3"

Insert length: 3584 bp
Poly A stretch at pos. 3555, polyadenylation signal at pos. 3537

```
   1 GGGCGGGAGG TGCTTCCCAA GGACCGTAGA TGCCTCTCTA GAGCATGAGC
  51 TCAGGCAAGA GTGCCCGCTA CAACCGCTTC TCCGGGGGGC CCAGCAATCT
 101 TCCCACCCCA GACGTCACCA CAGGGACCAG AATGGAAACG ACCTTCGGAC
 151 CCGCCTTTTC AGCCGTCACC ACCATCACAA AAGCTGACGG GACCAGCACC
 201 TACAAGCAGC ACTGCAGGAC AGCATGCCCC CCATCAGCAC TCCCCGCCGC
 251 TCCGACTCCG CCATCTCTGT CCGCTCCCTG CACTCAGAGT CCAGCATGTC
 301 TCTGCGCTCC ACATTCTCAC TGCCCGAGGA GGAGGAGGAG CCGGAGCCAC
 351 TGGTGTTTGC GGAGCAGCCC TCGGTGAAGC TGTGCTGTCA GCTCTGCTGC
 401 AGCGTCTTCA AAGACCCCGT GATCACCACG TGTGGGCACA CGTTCTGTAG
 451 GAGATGCGCC TTGAAGTCAG AGAAGTGTCC CGTGGACAAC GTCAAACTGA
 501 CCGTGGTGGT GAACAACATC GCGGTGGCCG AGCAGATCGG GGAGCTCTTC
 551 ATCCACTGCC GGCACGGCTG CCGGGTAGCG GGCAGCGGGA AGCCCCCCAT
 601 CTTTGAGGTG GACCCCCGAG GGTGCCCCTT CACCATCAAG CTCAGCGCCC
 651 GGAAGGACCA CGAGGGCAGC TGTGACTACA GGCCTGTGCG GTGTCCCAAC
 701 AACCCCAGCT GCCCCCCGCT GCTCAGGATG AACCTGGAGG CCCACCTCAA
 751 GGAGTGCGAG CACATCAAAT GCCCCCACTC CAAGTACGGG TGCACGTTCA
 801 TCGGGAACCA GGACACTTAC GAGACCCACC TGGAGACTTG CCGCTTCGAG
 851 GGCCTGAAGG AGTTTCTGCA GCAGACGGAT GACCGCTTCC ACGAGATGCA
 901 CGTGGCTCTG GCCCAGAAGG ACCAGGAGAT CGCCTTCCTG CGCTCCATGC
 951 TGGGAAAGCT CTCGGAGAAG ATCGACCAGC TAGAGAAGAG CCTGGAGCTC
1001 AAGTTTGACG TCCTGGACGA AAACCAGAGC AAGCTCAGCG AGGACCTCAT
1051 GGAGTTCCGG CGGGACGCAT CCATGTTAAA TGACGAGCTG TCCCACATCA
1101 ACGCGCGGCT GAACATGGGC ATCCTAGGCT CCTACGACCC TCAGCAGATC
1151 TTCAAGTGCA AAGGGACCTT TGTGGGCCAC CAGGGCCCTG TGTGGTGTCT
1201 CTGCGTCTAC TCCATGGGTG ACCTGCTCTT CAGTGGCTCC TCTGACAAGA
1251 CCATCAAGGT GTGGGACACA TGTACCACCT ACAAGTGTCA GAAGACACTG
1301 GAGGGCCATG ATGGCATCGT GCTGGCTCTC TGCATCCAGG GGTGCAAACT
1351 CTACAGCGGC TCTGCAGACT GCACCATCAT TGTGTGGGAC ATCCAGAACC
1401 TGCAGAAGGT GAACACCATC CGGGCCCATG ACAACCCGGT GTGCACGCTG
1451 GTCTCCTCAC ACAACGTGCT CTTCAGCGGC TCCCTGAAGG CCATCAAGGT
1501 CTGGGACATC GTGGGCACTG AGCTGAAGTT GAAGAAGGAG CTCACAGGCC
1551 TCAACCACTG GGTGCGGGCC CTGGTGGCTG CCCAGAGCTA CCTGTACAGC
1601 GGCTCCTACC AGACAATCAA GATCTGGGAC ATCCGAACCC TTGACTGCAT
1651 CCACGTCCTG CAGACGTCTG GTGGCAGCGT CTACTCCATT GCTGTGACAA
1701 ATCACCACAT TGTCTGTGGC ACCTACGAGA ACCTCATCCA CGTGTGGGAC
1751 ATTGAGTCCA AGGAGCAGGT GCGGACCCTC ACGGGCCACG TGGGCACCGT
1801 GTATGCCCTG GCGGTCATCT CGACGCCAGA CCAGACCAAA GTCTTCAGTG
1851 CATCCTACGA CCGGTCCCTC AGGGTCTGGA GTATGGACAA CATGATCTGC
1901 ACGCAGACCC TGCTGCGTCA CCAGGGCAGT GTCACCGCGC TGGCTGTGTC
1951 CCGGGGCCGA CTCTTCTCAG GGCTGTGGA TAGCACTGTG AAGGTTTGGA
2001 CTTGCTAACA GGATCCAGGC CAGGCTGTGG TTTCCCCTGA ACCAGCCCTG
2051 GACCTTTCTG AGCCAGGCTG GCCACATGGG GTGGTCTCGG GGTTTCTGCC
2101 TGCCCCGTGG GCATAGGTGG ACAGGCTCTG GCAGCCGGGC AGTGCCCTCC
2151 CCGTCCCATG CTCGGCGAGC CTCCCTCTAC TCGGCACTGT CCTTGCTGCC
2201 CAGCCCCTCT CTGGGTGCCA GGTACGACGC TTGCCCCGGC CCACCCTCCA
2251 TCCCCACCCT CCATCCCCAC CCTAGATGGA GCGAGGGCCT TTTTACTCAC
2301 CTTTTCTACC GTTTTTAGAC TGTATGTAGA TTTGGTTACC TCCTGGTTGA
```

```
2351 AATAAATGCT CCACAGACTG TGGCTGTGAG TGGGGACAGC TCCTCGGGAC
2401 AAGGGGGCTG TGTGTGGCCT TGAGGTTGGT GTGCACAGGC ACTGGCTGCT
2451 GTGAGTGGGG GGGCATGGGG CAGTTTCCTT TGGTGGACCC CAGGACTTCG
2501 GCCCACTCCG GGGCCTCCCC TCCCTGCTAG GAGGCAACTC GTCACACCCA
2551 AGCTGCTGGC CTCCAGTCCC ATCTCCCCCA ACACATGTGC CCCCAAAAAG
2601 TGAGCCAGGC ACCTCTGTTT CCTGCTGTTT ATTGACAGCC GACGGCAGCG
2651 CCTTGCCCAG ACCTCCCCTG CCCACCTGCT GGAGCCCAGC CTGTGCCGCC
2701 CTCTGAGGAG AGGCCTGGGG GGACAGCTGG GCACGTCCAC TCGCAGGGAA
2751 ACACGGGGTG AGACAGCAGG AAGGGGCCCT GCACGCCGGG ACGCCACCTC
2801 CGCCAGCCGC CTCCACCCGC CCCACACCAC AATCGCTGGT TTTCGGCATT
2851 TTTTAAATTT TTTTTTTAAG AAACGTCAAA GTTGTGCCCA ACACTGTGGA
2901 TCAGCAAACA CGATAGAGGA GACCAGTCAG TACTTCTTGG AGGGGGCAGG
2951 AGGAGAGAGG AAAAGGGAGG GCGAGAATGA CCACACAACA CAGCCTTGGA
3001 CCATGAGCAG AAGCGTCCGT GGGAACTCCA CTGGGGTGGA TGGGCTGCCT
3051 GCACAGCCCC TGGAGAGGGG GCCAGGCACA CCCTCAGAGG AGCTGCAAGC
3101 CCGTGGCCTG GCCTGCTACA TGCCCTGCTT CCACGTGGCT GCCACGCTGA
3151 CACACCCACA TTCACCAAAC CCACCCGCGC CCTGGGACGC AGCCACGCCA
3201 GGAGGAGGAC ACGGCCGCCG AGAGCAAGGC ACAACCTCGA GTTCTTGGGG
3251 CGCAGAGAAC TTAGGAGAGA AGCACGGAGG AGCCCCCGGC AGAGCACCCG
3301 CCCCCGGGCC CCAGCCTTCC ACCTGTGCTA GCAGCCTGGG GCCTCCACTC
3351 TGGCCGGAGG AAGGACCGCA GGCAGACAGC CTGGGCCTCT AACAGCTTTT
3401 GTCCGGAGCT AGACTTCGTG TCCTTTCAGT TGGTAAATGG TTTTCTATAG
3451 AATCAATAAT ATTTCTTTCT TTAAATATAT ATTTGTTAAA GTTATACCTT
3501 TTTGTTTCTC TGGGGAAATC CGCCTCAGCT CATTCCCAAT AAATTAATAC
3551 TCTTGATAAA AAAAAAAAAA AGAAAAAAAA AAAA
```

BLAST Results

Entry HSBE from database EMBL:
Homo sapiens (clone exon trap d5) chromosome 16p13.3 gene, exon.
Score = 2375, P = 7.1e-101, identities = 475/475

Entry HSBD from database EMBL:
Homo sapiens (clone exon trap d32) chromosome 16p13.3 gene, exon.
Score = 876, P = 3.0e-31, identities = 176/177

Medline entries

95122486:
Structural analysis of myosin heavy chain kinase A from Dictyostelium. Evidence for a highly divergent protein kinase domain, an amino-terminal coiled-coil domain, and a domain homologous to the beta-subunit of heterotrimeric G proteins.

96149460:
Dictyostelium myosin heavy chain kinase A regulates myosin localization during growth and development.

97277316:
Identification of a protein kinase from Dictyostelium with homology to the novel catalytic domain of myosin heavy chain kinase A.

96009891:
A gene responsible for vegetative incompatibility in the fungus Podospora anserina encodes a protein with a GTP-binding motif and G beta homologous domain.

Peptide information for frame 2

ORF from 224 bp to 2005 bp; peptide length: 594
Category: similarity to known protein
Prosite motifs: ZINC_FINGER_C3HC4 (70-80)
LEUCINE_ZIPPER (436-458)
LEUCINE_ZIPPER (436-458)
G_BETA_REPEATS (335-355)
G_BETA_REPEATS (376-391)

```
  1 MPPISTPRRS DSAISVRSLH SESSMSLRST FSLPEEEEEP EPLVFAEQPS
 51 VKLCCQLCCS VFKDPVITTC GHTFCRRCAL KSEKCPVDNV KLTVVVNNIA
101 VAEQIGELFI HCRHGCRVAG SGKPPIFEVD PRGCPFTIKL SARKDHEGSC
151 DYRPVRCPNN PSCPPLLRMN LEAHLKECEH IKCPHSKYGC TFIGNQDTYE
201 THLETCRFEG LKEFLQQTDD RFHEMHVALA QKDQEIAFLR SMLGKLSEKI
251 DQLEKSLELK FDVLDENQSK LSEDLMEFRR DASMLNDELS HINARLNMGI
301 LGSYDPQQIF KCKGTFVGHQ GPVWCLCVYS MGDLLFSGSS DKTIKVWDTC
351 TTYKCQKTLE GHDGIVLALC IQGCKLYSGS ADCTIIVWDI QNLQKVNTIR
401 AHDNPVCTLV SSHNVLFSGS LKAIKVWDIV GTELKLKKEL TGLNHWVRAL
451 VAAQSYLYSG SYQTIKIWDI RTLDCIHVLQ TSGGSVYSIA VTNHHIVCGT
501 YENLIHVWDI ESKEQVRTLT GHVGTVYALA VISTPDQTKV FSASYDRSLR
551 VWSMDNMICT QTLLRHQGSV TALAVSRGRL FSGAVDSTVK VWTC
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_1i2, frame 2

SWISSPROT:KMHB_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B)., N = 1, Score = 419, P = 3.6e-37

SWISSPROT:HET1_PODAN VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1., N = 1, Score = 392, P = 3.1e-33

SWISSPROT:YDJ5_SCHPO HYPOTHETICAL 67.1 KD TRP-ASP REPEATS CONTAINING PROTEIN C57A10.05C IN CHROMOSOME I., N = 1, Score = 357, P = 4.1e-30

TREMBL:AF032878_1 gene: "slimb"; product: "Slimb"; Drosophila melanogaster Slimb (slimb) mRNA, complete cds., N = 1, Score = 347, P = 1.7e-29

>SWISSPROT:KMHB_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B).
Length = 732

HSPs:

```
 Score = 419 (62.9 bits), Expect = 3.6e-37, P = 3.6e-37
 Identities = 96/268 (35%), Positives = 158/268 (58%)

Query:   325 CLCVYSMGDLLFSGSSDKTIKVWD-TCTTYKCQKTLEGHDGIVLALCIQGCKLYSGSADC 383
             C+C     +LLF+G SD +I+V+D      +C +TL+GH+G V ++C     L+SGS+D
Sbjct:   467 CIC----DNLLFTGCSDNSIRVYDYKSQNMECVQTLKGHEGPVESICYNDQYLFSGSSDH 522

Query:   384 TIIVWDIQNLQKVNTIRAHDNPVCTLVSSHNVLFSGSL-KAIKVWDIVGTELKLKKELTG 442
             +I VWD++ L+ + T+  HD PV T++ +   LFSGS  K IKVWD+    L+ K  L
Sbjct:   523 SIKVWDLKKLRCIFTLEGHDKPVHTVLLNDKYLFSGSSDKTIKVWDL--KTLECKYTLES 580

Query:   443 LNHWVRALVAAQSYLYSGSY-QTIKIWDIRTLDCIHVLQTSGGSVYSIAVTNHHIVCGTY 501
                 V+ L  +  YL+SGS  +TIK+WD++T  C + L+     V +I +   ++  G+Y
Sbjct:   581 HARAVKTLCISGQYLFSGSNDKTIKVWDLKTFRCNYTLKGHTKWVTTICILGTNLYSGSY 640

Query:   502 ENLIHVWDIESKEQVRTLTGHVGTVYALAVISTPDQTKVFSASYDRSLRVWSMDNMICTQ 561
             +  I VW+++S E   TL GH   V + +     D+  +F+AS D ++++W ++ + C
Sbjct:   641 DKTIRVWNLKSLECSATLRGHDRWVEHMVIC---DKL-LFTASDDNTIKIWDLETLRCNT 696

Query:   562 TLLRHQGSVTALAVSRGR--LFSGAVDSTVKVW 592
             TL  H  +V  LAV   +  + S + D +++VW
Sbjct:   697 TLEGHNATVQCLAVWEDKKCVISCSHDQSIRVW 729

 Score = 415 (62.3 bits), Expect = 1.2e-36, P = 1.2e-36
 Identities = 113/303 (37%), Positives = 166/303 (54%)

Query:   255 KSLEL-KFDVLDENQSKLSEDLMEFRRDASMLNDEL-SHINARLNMGILGS-------YD 305
             KS++L K ++L  N+ K S +L   +   ++    + SH+    N+   G        YD
Sbjct:   427 KSIDLEKPEILINNKKKESINLETIKLIETIKGYHVTSHLCICDNLLFTGCSDNSIRVYD 486

Query:   306 -PQQIFKCKGTFVGHQGPVWCLCVYSMGDLLFSGSSDKTIKVWDTCTTYKCQKTLEGHDG 364
                Q  +C  T  GH+GPV  +C Y+    LFSGSSD +IKVWD     +C  TLEGHD
Sbjct:   487 YKSQNMECVQTLKGHEGPVESIC-YN-DQYLFSGSSDHSIKVWDL-KKLRCIFTLEGHDK 543

Query:   365 IVLALCIQGCKLYSGSADCTIIVWDIQNLQKVNTIRAHDNPVCTLVSSHNVLFSGSL-KA 423
              V + +     L+SGS+D TI VWD++ L+   T+ +H   V TL  S   LFSGS  K
Sbjct:   544 PVHTVLLNDKYLFSGSSDKTIKVWDLKTLECKYTLESHARAVKTLCISGQYLFSGSNDKT 603

Query:   424 IKVWDIVGTELKLKKELTGLNHWVRALVAAQSYLYSGSY-QTIKIWDIRTLDCIHVLQTS 482
             IKVWD+      +    L G   WV  +    + LYSGSY +TI++W++++L+C   L+
Sbjct:   604 IKVWDL--KTFRCNYTLKGHTKWVTTICILGTNLYSGSYDKTIRVWNLKSLECSATLRGH 661
```

```
Query:   483 GGSVYSIAVTNHHIVCGTYENLIHVWDIESKEQVRTLTGHVGTVYALAVISTPDQTKVFS 542
             V  + + +   +    + +N I +WD+E+      TL GH  TV  LAV      D+  V S
Sbjct:   662 DRWVEHMVICDKLLFTASDDNTIKIWDLETLRCNTTLEGHNATVQCLAVWE--DKKCVIS 719

Query:   543 ASYDRSLRVW 552
             S+D+S+RVW
Sbjct:   720 CSHDQSIRVW 729

 Score = 262 (39.3 bits), Expect = 3.2e-19, P = 3.2e-19
 Identities = 60/184 (32%), Positives = 109/184 (59%)

Query:   352 TYKCQKTLEGHDGIVLALCIQGCKLYSGSADCTIIVWDI--QNLQKVNTIRAHDNPVCTL 409
             T K  +T++G+  +    LCI     L++G +D +I V+D    QN++  V T++ H+  PV ++
Sbjct:   450 TIKLIETIKGYH-VTSHLCICDNLLFTGCSDNSIRVYDYKSQNMECVQTLKGHEGPVESI 508

Query:   410 VSSHNVLFSGSLK-AIKVWDIVGTELKLKKELTGLNHWVRALVAAQSYLYSGSY-QTIKI 467
              +     LFSGS    +IKVWD+     +L+      L G +  V   ++      YL+SGS   +TIK+
Sbjct:   509 CYNDQYLFSGSSDHSIKVWDL--KKLRCIFTLEGHDKPVHTVLLNDKYLFSGSSDKTIKV 566

Query:   468 WDIRTLDCIHVLQTSGGSVYSIAVTNHHIVCGTYENLIHVWDIESKEQVRTLTGHVGTVY 527
             WD++TL+C + L++    +V ++ ++     ++  G+ +  I VWD+++        TL GH    V
Sbjct:   567 WDLKTLECKYTLESHARAVKTLCISGQYLFSGSNDKTIKVWDLKTFRCNYTLKGHTKWVT 626

Query:   528 ALAVIST 534
             + ++ T
Sbjct:   627 TICILGT 633

 Score = 173 (26.0 bits), Expect = 1.7e-09, P = 1.7e-09
 Identities = 43/118 (36%), Positives = 65/118 (55%)

Query:   310 FKCKGTFVGHQGPVWCLCVYSMGDLLFSGSSDKTIKVWDTCTTYKCQKTLEGHDGIVLAL 369
             F+C   T  GH    V   +C+   +G   L+SGS DKTI+VW+    + +C  TL GHD  V  +
Sbjct:   612 FRCNYTLKGHTKWVTTICI--LGTNLYSGSYDKTIRVWNL-KSLECSATLRGHDRWVEHM 668

Query:   370 CIQGCKLYSGSADCTIIVWDIQNLQKVNTIRAHDNPV-CTLVSSHN--VLFSGSLKAIKV 426
             I    L++ S D TI +WD++ L+    T+ H+  V C V    V+       ++I+V
Sbjct:   669 VICDKLLFTASDDNTIKIWDLETLRCNTTLEGHNATVQCLAVWEDKKCVISCSHDQSIRV 728

Query:   427 W 427
             W
Sbjct:   729 W 729
```

Pedant information for DKFZphutel_1i2, frame 2

Report for DKFZphutel_1i2.2

```
[LENGTH]      594
[MW]          66541.94
[pI]          6.64
[HOMOL]       SWISSPROT:KMHB_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B). 3e-37

[FUNCAT]      03.22 cell cycle control and mitosis  [S. cerevisiae, YIL046w] 5e-21
[FUNCAT]      06.13.01 cytoplasmic degradation      [S. cerevisiae, YIL046w] 5e-21
[FUNCAT]      04.05.01.04 transcriptional control  [S. cerevisiae, YIL046w] 5e-21
[FUNCAT]      30.10 nuclear organization   [S. cerevisiae, YIL046w] 5e-21
[FUNCAT]      01.01.04 regulation of amino-acid metabolism        [S. cerevisiae, YIL046w]
5e-21
[FUNCAT]      99 unclassified proteins      [S. cerevisiae, YCR072c beta-transducin family]
2e-15
[FUNCAT]      30.04 organization of cytoskeleton    [S. cerevisiae, YFL009w] 1e-14
[FUNCAT]      03.04 budding, cell polarity and filament formation  [S. cerevisiae, YFL009w]
1e-14
[FUNCAT]      03.10 sporulation and germination     [S. cerevisiae, YFL009w] 1e-14
[FUNCAT]      03.16 dna synthesis and replication   [S. cerevisiae, YFL009w] 1e-14
[FUNCAT]      30.09 organization of intracellular transport vesicles      [S. cerevisiae,
YDL145c] 1e-13
[FUNCAT]      08.07 vesicular transport (golgi network, etc.)      [S. cerevisiae, YDL145c]
1e-13
[FUNCAT]      04.05.03 mrna processing (splicing)   [S. cerevisiae, YPR178w] 2e-11
[FUNCAT]      06.10 assembly of protein complexes   [S. cerevisiae, YPR178w] 2e-11
[FUNCAT]      04.05.01.01 general transcription activities        [S. cerevisiae, YBR198c
TAF90 - TFIID subunit] 3e-11
[FUNCAT]      03.13 meiosis  [S. cerevisiae, YLR129w] 8e-09
[FUNCAT]      30.03 organization of cytoplasm       [S. cerevisiae, YCR057c] 2e-07
[FUNCAT]      03.25 cytokinesis       [S. cerevisiae, YCR057c] 2e-07
[FUNCAT]      02.16 fermentation      [S. cerevisiae, YMR116c] 5e-07
[FUNCAT]      05.04 translation (initiation, elongation and termination)  [S. cerevisiae,
YMR116c] 5e-07
```

```
[FUNCAT]        06.13 proteolysis      [S. cerevisiae, YGL003c] 3e-06
[FUNCAT]        03.01 cell growth      [S. cerevisiae, YKL021c] 2e-04
[FUNCAT]        01.03.07 deoxyribonucleotide metabolism      [S. cerevisiae, YOR269w] 2e-04
[FUNCAT]        30.02 organization of plasma membrane      [S. cerevisiae, YOR212w] 0.001
[FUNCAT]        10.05.07 g-proteins    [S. cerevisiae, YOR212w] 0.001
[FUNCAT]        03.07 pheromone response, mating-type determination, sex-specific proteins
      [S. cerevisiae, YOR212w] 0.001
[BLOCKS]        BL00678
[BLOCKS]        BL00518 Zinc finger, C3HC4 type, proteins
[SCOP]          d1tbgd_ 2.46.3.1.1 beta1-subunit of the signal-transducing  3e-10
[EC]            2.7.1.129 Myosin-heavy-chain kinase 3e-26
[PIRKW]         phosphotransferase 3e-26
[PIRKW]         nucleus 1e-06
[PIRKW]         plasma 9e-08
[PIRKW]         duplication 3e-25
[PIRKW]         hormone 9e-08
[PIRKW]         zinc 3e-09
[PIRKW]         cell cycle control 4e-13
[PIRKW]         transmembrane protein 3e-12
[PIRKW]         zinc finger 1e-08
[PIRKW]         stomach 9e-08
[PIRKW]         DNA binding 9e-06
[PIRKW]         autophosphorylation 3e-26
[PIRKW]         phosphoprotein 3e-26
[PIRKW]         signal transduction 5e-08
[PIRKW]         heterotrimer 5e-08
[PIRKW]         coiled coil 3e-26
[PIRKW]         multimer 3e-26
[PIRKW]         transcription regulation 4e-10
[PIRKW]         GTP binding 5e-08
[SUPFAM]        chromobox homology 9e-06
[SUPFAM]        RING finger homology 3e-09
[SUPFAM]        coatomer complex beta' chain 1e-07
[SUPFAM]        WD repeat homology 3e-26
[SUPFAM]        yeast coatomer complex alpha chain 3e-12
[SUPFAM]        GTP-binding regulatory protein beta chain 5e-08
[SUPFAM]        PRL1 protein 2e-09
[PROSITE]       WD_REPEATS      2
[PROSITE]       LEUCINE_ZIPPER 1
[PROSITE]       MYRISTYL        14
[PROSITE]       CK2_PHOSPHO_SITE        4
[PROSITE]       ZINC_FINGER_C3HC4       1
[PROSITE]       PKC_PHOSPHO_SITE        18
[PROSITE]       ASN_GLYCOSYLATION       1
[PFAM]          Zinc finger, C3HC4 type (RING finger)
[PFAM]          WD domain, G-beta repeats
[KW]            Irregular
[KW]            3D
[KW]            LOW_COMPLEXITY      6.23 %
[KW]            COILED_COIL         6.73 %
```

```
SEQ     MPPISTPRRSDSAISVRSLHSESSMSLRSTFSLPEEEEEPEPLVFAEQPSVKLCCQLCCS
SEG     ..............xxxxxxxxxxxxxxx....xxxxxxxxx..................
COILS   ............................................................
1gg2B   ............................................................

SEQ     VFKDPVITTCGHTFCRRCALKSEKCPVDNVKLTVVVNNIAVAEQIGELFIHCRHGCRVAG
SEG     ............................................................
COILS   ............................................................
1gg2B   ............................................................

SEQ     SGKPPIFEVDPRGCPFTIKLSARKDHEGSCDYRPVRCPNNPSCPPLLRMNLEAHLKECEH
SEG     ............................................................
COILS   ............................................................
1gg2B   ............................................................

SEQ     IKCPHSKYGCTFIGNQDTYETHLETCRFEGLKEFLQQTDDRFHEMHVALAQKDQEIAFLR
SEG     ............................................................
COILS   ..............................................CCCCCCCCCCCCCC
1gg2B   ............................................................

SEQ     SMLGKLSEKIDQLEKSLELKFDVLDENQSKLSEDLMEFRRDASMLNDELSHINARLNMGI
SEG     ............................................................
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCC..................................
1gg2B   ............................................................

SEQ     LGSYDPQQIFKCKGTFVGHQGPVWCLCVYSMGDLLFSGSSDKTIKVWDTCTTYKCQKTLE
SEG     ............................................................
COILS   ............................................................
1gg2B   ..............EECCCCCCEEEEEETTTTCEEEEEETTTEEEEEEG-GGCEEEEEEEE
```

```
SEQ    GHDGIVLALCIQGCKLYSGSADCTIIVWDIQNLQKVNTIRAHDNPVCTLVSSHNVLFSGS
SEG    ............................................................
COILS  ............................................................
1gg2B  CCCCCEEEEEETTCEEEEEETTTCEEEEETTTTEEEEEE-CTTTTCCCEEE.........

SEQ    LKAIKVWDIVGTELKLKKELTGLNHWVRALVAAQSYLYSGSYQTIKIWDIRTLDCIHVLQ
SEG    ..........xxxxxxxxxxxxx.....................................
COILS  ............................................................
1gg2B  ............................................................

SEQ    TSGGSVYSIAVTNHHIVCGTYENLIHVWDIESKEQVRTLTGHVGTVYALAVISTPDQTKV
SEG    ............................................................
COILS  ............................................................
1gg2B  ............................................................

SEQ    FSASYDRSLRVWSMDNMICTQTLLRHQGSVTALAVSRGRLFSGAVDSTVKVWTC
SEG    ......................................................
COILS  ......................................................
1gg2B  ......................................................
```

Prosite for DKFZphutel_1i2.2

PS00001	267->271	ASN_GLYCOSYLATION	PDOC00001
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	15->18	PKC_PHOSPHO_SITE	PDOC00005
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	50->53	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	121->124	PKC_PHOSPHO_SITE	PDOC00005
PS00005	137->140	PKC_PHOSPHO_SITE	PDOC00005
PS00005	141->144	PKC_PHOSPHO_SITE	PDOC00005
PS00005	205->208	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	340->343	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	352->355	PKC_PHOSPHO_SITE	PDOC00005
PS00005	398->401	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	464->467	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	588->591	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	330->334	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00008	115->121	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	194->200	MYRISTYL	PDOC00008
PS00008	299->305	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	364->370	MYRISTYL	PDOC00008
PS00008	379->385	MYRISTYL	PDOC00008
PS00008	419->425	MYRISTYL	PDOC00008
PS00008	460->466	MYRISTYL	PDOC00008
PS00008	484->490	MYRISTYL	PDOC00008
PS00008	499->505	MYRISTYL	PDOC00008
PS00008	524->530	MYRISTYL	PDOC00008
PS00008	568->574	MYRISTYL	PDOC00008
PS00008	583->589	MYRISTYL	PDOC00008
PS00518	70->80	ZINC_FINGER_C3HC4	PDOC00449
PS00029	436->458	LEUCINE_ZIPPER	PDOC00029
PS00678	335->350	WD_REPEATS	PDOC00574
PS00678	376->391	WD_REPEATS	PDOC00574

Pfam for DKFZphutel_1i2.2

HMM_NAME WD domain, G-beta repeats

```
HMM              *MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*
                  ++GH ++VWC+ +    G + ++SGS D+T+++WD
Query       316   FVGHQGPVWCLCVYSMGDL-LFSGSSDKTIKVWD    348
```

22.93 519 553 1 34 dkfzphutel_1i2.2 similarity to Dictostelium myosin heavy chain kinase

Alignment to HMM consensus:

```
Query                  *MrGHnnWVWCVaF..SPDGrWFIvSGSWDgTCRLWD*
                        ++GH ++V+++A+  +PD    ++S+S D+++R+W+
  dkfzphutel    519     LTGHVGTVYALAVISTPDQTK-VFSASYDRSLRVWS    553


HMM_NAME        Zinc finger, C3HC4 type (RING finger)

HMM                    *CPICFcTFQlDyPWPFdePmMlPCgHsFCypCIrrW..CPmC*
                        C++C      +      F++P++++CGH+FC+ C +++  CP+
Query            55     CQLC-----CSV---FKDPVITTCGHTFCRRCALKSEKCPVD      88
```

DKFZphutel_20b19

group: metabolism

DKFZphutel_20b19 encodes a novel 486 amino acid protein with similarity to bacterial sarcosine oxidases (EC 1.5.3.1.)

The novel protein seems to be a novel enzyme with sarcosine oxidase activity.

The new protein can find application in modulation of sarcosine metabolism and as a new enzyme for biotechnologic production processes.

similarity to sarcosine oxidases
membrane regions: 1
Summary DKFZphutel_20b19 encodes a novel 486 amino acid protein, with similarity to sarcosine oxidases.

similarity to sarcosine oxidases

complete cDNA?, complete cds potential start at Bp 48, EST hits,

Sequenced by AGOWA

Locus: unknown

Insert length: 1967 bp
Poly A stretch at pos. 1950, no polyadenylation signal found

```
   1 AGCGAGGCAG CAGTGCAGCT TTCAGAGGGT CCGGGCTCAG AGGGGTTATG
  51 ATTCGGAGGG TTCTGCCGCA CGGCATGGGC CGGGGCCTCT TGACCCGGAG
 101 GCCAGGCACG CGCAGAGGAG GCTTTTCTCT GGACTGGGAT GGAAAGGTGT
 151 CTGAGATTAA GAAGAAGATC AAGTCGATCC TGCCTGGAAG GTCCTGTGAT
 201 CTACTGCAAG ACACCAGCCA CCTGCCTCCC GAGCACTCGG ATGTGGTGAT
 251 CGTGGGAGGT GGGGTGCTTG GCTTGTCTGT GGCCTATTGG CTGAAGAAGC
 301 TGGAGAGCAG ACGAGGTGCT ATTCGAGTGC TAGTGGTGGA ACGGGACCAC
 351 ACGTATTCAC AGGCCTCCAC TGGGCTCTCA GTAGGTGGGA TTTGTCAGCA
 401 GTTCTCATTG CCTGAGAACA TCCAGCTCTC CCTCTTTTCA GCCAGCTTTC
 451 TACGGAACAT CAATGAGTAC CTGGCCGTAG TCGATGCTCC TCCCCTGGAC
 501 CTCCGGTTCA ACCCCTCGGG CTACCTCTTG CTGGCTTCAG AAAAGGATGC
 551 TGCAGCCATG GAGAGCAACG TGAAAGTGCA GAGGCAGGAG GGAGCCAAAG
 601 TTTCTCTGAT GTCTCCTGAT CAGCTTCGGA ACAAGTTTCC CTGGATAAAC
 651 ACAGAGGGAG TGGCTTTGGC GTCTTATGGG ATGGAGGACG AAGGTTGGTT
 701 TGACCCCTGG TGTCTGCTCC AGGGGCTTCG GCGAAAGGTC CAGTCCTTGG
 751 GAGTCCTTTT CTGCCAGGGA GAGGTGACAC GTTTTGTCTC TTCATCTCAA
 801 CGCATGTTGA CCACAGATGA CAAAGCGGTG GTCTTGAAAA GGATCCATGA
 851 AGTCCATGTG AAGATGGACC GCAGCCTGGA GTACCAGCCT GTGGAATGCG
 901 CCATTGTGAT CAACGCAGCC GGAGCCTGGT CTGCGCAAAT CGCAGCACTG
 951 GCTGGTGTTG GAGAGGGGCC GCCTGGCACC CTGCAGGGCA CCAAGCTACC
1001 TGTGGAGCCG AGGAAAAGGT ATGTGTATGT GTGGCACTGC CCCCAGGGAC
1051 CAGGCCTAGA GACTCCGCTT GTTGCAGACA CCAGTGGAGC CTATTTTCGC
1101 CGGGAAGGAT TAGGTAGCAA CTACCTAGGT GGTCGTAGCC CCACTGAGCA
1151 GGAAGAACCG GACCCGGCGA ACCTGGAAGT GGACCATGAT TTCTTCCAGG
1201 ACAAGGTGTG GCCCCATTTG GCCCTGAGGG TCCCAGCTTT TGAGACTCTG
1251 AAGGTTCAGA GCGCCTGGGC CGGCTATTAC GACTACAACA CCTTTGACCA
1301 GAATGGCGTG GTGGGCCCCC ACCCGCTAGT TGTCAACATG TACTTTGCTA
1351 CTGGCTTCAG TGGTCACGGG CTCCAGCAGG CCCCTGGCAT TGGGCGAGCT
1401 GTAGCAGAGA TGGTACTGAA GGGCAGGTTC CAGACCATCG ACCTGAGCCC
1451 CTTCCTCTTT ACCCGCTTTT ACTTGGGAGA GAAGATCCAG GAGAACAACA
1501 TCATCTGAGC ATGTGTGCTC TGCACTGGCT CCACTGGCTT GCATCCTGGC
1551 TGTGTTCACA GCCTTGTTTG CTGCTTCCAT CTTCCCCAGT ACTGTGCCAG
1601 GCCTTCTCCC CCTCCCCAGT GTCCTCTCCT CTCAGGCAGG CCATTGCACC
1651 CATATGGCTG GGCAGGCACA GGCAGTGAGG CCGAGGCCAA TAGCGAGTGA
1701 TGAGCGGGAT CCTAGGACTG ATCTGTAGCC CATGCTGATG TCACCCACCA
1751 GGGCAATCCA TCTGGAGGCC TGAGCACCCT GGCCCAGGAC TGGCTTCATC
1801 CTGGCACTGA CCAGGAAAGA CTGCCTCTGA CCCTCTTAGC AGACAGAGCC
1851 CAGGCATGGG AGCACTCTGG GGCAGCCTGG CTCAGGTTTA TTGATTTTCG
1901 TCTGTTTACC CTATCCATTA ATCAATACAT GTAATTAACT CCTTCCCTCC
1951 AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 48 bp to 1505 bp; peptide length: 486
Category: similarity to known protein

```
  1 MIRRVLPHGM GRGLLTRRPG TRRGGFSLDW DGKVSEIKKK IKSILPGRSC
 51 DLLQDTSHLP PEHSDVVIVG GGVLGLSVAY WLKKLESRRG AIRVLVVERD
101 HTYSQASTGL SVGGICQQFS LPENIQLSLF SASFLRNINE YLAVVDAPPL
151 DLRFNPSGYL LLASEKDAAA MESNVKVQRQ EGAKVSLMSP DQLRNKFPWI
201 NTEGVALASY GMEDEGWFDP WCLLQGLRRK VQSLGVLFCQ GEVTRFVSSS
251 QRMLTTDDKA VVLKRIHEVH VKMDRSLEYQ PVECAIVINA AGAWSAQIAA
301 LAGVGEGPPG TLQGTKLPVE PRKRYVYVWH CPQGPGLETP LVADTSGAYF
351 RREGLGSNYL GGRSPTEQEE PDPANLEVDH DFFQDKVWPH LALRVPAFET
401 LKVQSAWAGY YDYNTFDQNG VVGPHPLVVN MYFATGFSGH GLQQAPGIGR
451 AVAEMVLKGR FQTIDLSPFL FTRFYLGEKI QENNII
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_20b19, frame 3

TREMBL:CEM04B2_4 gene: "M04B2.4"; Caenorhabditis elegans cosmid M04B2, N = 1, Score = 801, P = 9.2e-80

PIR:B71184 probable sarcosine oxidase - Pyrococcus horikoshii, N = 2, Score = 194, P = 2e-26

PIR:B69284 sarcosine oxidase, subunit beta (soxB) homolog - Archaeoglobus fulgidus, N = 3, Score = 189, P = 8.2e-22

TREMBL:AF042732_1 gene: "Bb"; product: "unknown protein"; Anopheles gambiae (Bb) gene, partial cds; and TU37B2 (TU37B2) and diphenol oxidase-A2 (Dox-A2) genes, complete cds., N = 1, Score = 386, P = 8.7e-36

PIR:F71008 probable sarcosine oxidase - Pyrococcus horikoshii, N = 2, Score = 200, P = 4e-25

>TREMBL:CEM04B2_4 gene: "M04B2.4"; Caenorhabditis elegans cosmid M04B2
Length = 527

HSPs:

Score = 801 (120.2 bits), Expect = 9.2e-80, P = 9.2e-80
Identities = 171/433 (39%), Positives = 260/433 (60%)

```
Query:    61 PEHSDVVIVGGGVLGLSVAYWLKKLESRRGAIRVLVVERDHTYSQASTGLSVGGICQQFS 120
             P  +++VI+GGG+ G S A+WLK+   R    +V+VVE +  ++++ST LS GGI QQFS
Sbjct:    91 PYRAEIVIIGGGLSGSSTAFWLKE-RFRDEDFKVVVVENNDVFTKSSTMLSTGGITQQFS 149

Query:   121 LPENIQLSLFSASFLRNINEYLAVVDAPPLDLRFNPSGYLLLA-SEKDAAAMESNVKVQR 179
             +PE + +SLF+  FLR+  E+L ++D+   D+ F P+GYL LA ++++   M S  KVQ
Sbjct:   150 IPEFVDMSLFTTEFLRHAGEHLRILDSEQPDINFFPTGYLRLAKTDEEVEMMRSAWKVQI 209

Query:   180 QEGAKVSLMSPDQLRNKFPWINTEGVALASYGMEDEGWFDPWCLLQGLRRKVQSLGVLFC 239
             + GAKV L+S D+L  ++P++N + V LAS G+E+EG  D W LL  +R K  +LGV +
Sbjct:   210 ERGAKVQLLSKDELTKRYPYMNVDDVLLASLGVENEGTIDTWQLLSAIREKNITLGVQYV 269

Query:   240 QGEVTRFVSSSQRM----------LTTDDKAVVLKRIHEVHVKMDRS-LEYQPVECAIVI 288
             +GEV  F     R            T D+  +  +RI  V V+   +    +P+   +++
Sbjct:   270 KGEVEGFQFERHRASSEVHAFGDDATADENKLRAQRISGVLVRPQMNDASARPIRAHLIV 329

Query:   289 NAAGAWSAQIAALAGVGEGPPGTLQGTKLPVEPRKRYVYVWHCPQGPGLETPLVADTS-G 347
             NAAG W+ Q+A +AG+G+G  G L    +P++PRKR V+V   P  P  + P + D S G
Sbjct:   330 NAAGPWAGQVAKMAGIGKGT-GLL-AVPVPIQPRKRDVFVIFAPDVPS-DLPFIIDPSTG 386

Query:   348 AYFRREGLGSNYLGGRSPTEQEEP--DPANLEVDHDFFQDKVWPHLALRVPAFETLKVQS 405
              + R+   G  +L GR+P+++E+   D +NL+VD+D F  K+WP L  RVP F+T KV+S
Sbjct:   387 VFCRQTDSGQTFLVGRTPSKEEDAKRDHSNLDVDYDDFYQKIWPVLVDRVPGFQTAKVKS 446
```

```
Query:   406 AWAGYYDYNTFDQNGVVGPHPLVVNMYFATGFSGHGLQQAPGIGRAVAEMVLKGRFQTID 465
             AW+GY D NTFD   V+G HPL  N++   GF   G+  +     RA AE +  G +  ++
Sbjct:   447 AWSGYQDINTFDDAPVIGEHPLYTNLHMMCGFGERGVMHSMAAARAYAERIFDGAYINVN 506

Query:   466 LSPFLFTRFYLGEKIQE 482
             L  F   R    + I E
Sbjct:   507 LRKFDMRRIVKMDPITE 523
```

Pedant information for DKFZphutel_20b19, frame 3

Report for DKFZphutel_20b19.3

```
[LENGTH]      486
[MW]          53811.85
[pI]          7.66
[HOMOL]       TREMBL:CEM04B2_4 gene: "M04B2.4"; Caenorhabditis elegans cosmid M04B2 1e-78
[FUNCAT]      c energy conversion   [H. influenzae, HI0499] 8e-05
[BLOCKS]      BL00677A D-amino acid oxidases proteins
[BLOCKS]      BL00623A GMC oxidoreductases proteins
[BLOCKS]      BL01304A
[EC]          1.5.99.2 Dimethylglycine dehydrogenase 2e-07
[PIRKW]       flavoprotein 2e-07
[PIRKW]       oxidoreductase 2e-07
[PROSITE]     MYRISTYL        12
[PROSITE]     CK2_PHOSPHO_SITE      5
[PROSITE]     GLYCOSAMINOGLYCAN     1
[PROSITE]     PKC_PHOSPHO_SITE      6
[KW]          TRANSMEMBRANE  1
[KW]          LOW_COMPLEXITY     7.00 %
```

```
SEQ   MIRRVLPHGMGRGLLTRRPGTRRGGFSLDWDGKVSEIKKKIKSILPGRSCDLLQDTSHLP
SEG   ..........xxxxxxxxxxxxxx...........xxxxxxxx................
PRD   ccceeeccccccceeeccccccccccccccccchhhhhhhhhcccccceeeccccccc
MEM   ............................................................

SEQ   PEHSDVVIVGGGVLGLSVAYWLKKLESRRGAIRVLVVERDHTYSQASTGLSVGGICQQFS
SEG   .....xxxxxxxxxxx............................................
PRD   cccceeeeecccccchhhhhhhhhhhhhhccccceeeeeccccccccccccccccceeeec
MEM   .....MMMMMMMMMMMMMMMMM.......................................

SEQ   LPENIQLSLFSASFLRNINEYLAVVDAPPLDLRFNPSGYLLLASEKDAAAMESNVKVQRQ
SEG   ............................................................
PRD   ccchhhhhhhhhhhhhhhhhhhhhhhhhcccccceeeccccceeeehhhhhhhhhhhhhhh
MEM   ............................................................

SEQ   EGAKVSLMSPDQLRNKFPWINTEGVALASYGMEDEGWFDPWCLLQGLRRKVQSLGVLFCQ
SEG   ............................................................
PRD   cccceeecccchhhhhhcccccccccccccccccccccccchhhhhhhhhhhheeeeec
MEM   ............................................................

SEQ   GEVTRFVSSSQRMLTTDDKAVVLKRIHEVHVKMDRSLEYQPVECAIVINAAGAWSAQIAA
SEG   ............................................................
PRD   ceeeeeccccccccccccchhhhhhhhheeeeccccccccccceeeeeeecccchhhhhhh
MEM   ............................................................

SEQ   LAGVGEGPPGTLQGTKLPVEPRKRYVYVWHCPQGPGLETPLVADTSGAYFRREGLGSNYL
SEG   ............................................................
PRD   hhcccccccccccccccccccceeeeeeeccccccccccceeeccccceeeeccccccee
MEM   ............................................................

SEQ   GGRSPTEQEEPDPANLEVDHDFFQDKVWPHLALRVPAFETLKVQSAWAGYYDYNTFDQNG
SEG   ............................................................
PRD   eccccccccccccccccccccchhhhhhhhhhhhhcchhhhhhhhhheeeeeccccccc
MEM   ............................................................

SEQ   VVGPHPLVVNMYFATGFSGHGLQQAPGIGRAVAEMVLKGRFQTIDLSPFLFTRFYLGEKI
SEG   ............................................................
PRD   ccccccccccceeeeccccccccccchhhhhhhhhhhhhhcccceeeeccccccccccccccc
MEM   ............................................................

SEQ   QENNII
SEG   ......
PRD   cccccc
MEM   ......
```

Prosite for DKFZphutel_20b19.3

PS00002	438->442	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	87->90	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00005	250->253	PKC_PHOSPHO_SITE	PDOC00005
PS00005	400->403	PKC_PHOSPHO_SITE	PDOC00005
PS00006	120->124	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00006	364->368	CK2_PHOSPHO_SITE	PDOC00006
PS00006	366->370	CK2_PHOSPHO_SITE	PDOC00006
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	75->81	MYRISTYL	PDOC00008
PS00008	109->115	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	235->241	MYRISTYL	PDOC00008
PS00008	292->298	MYRISTYL	PDOC00008
PS00008	310->316	MYRISTYL	PDOC00008
PS00008	354->360	MYRISTYL	PDOC00008
PS00008	447->453	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphutel_20b19.3)

DKFZphutel_20g21
----

group: signal transduction

DKFZphutel_20g21 encodes a novel 861 amino acid protein with partial similarity to human ras inhibitor and other ras inhibitor proteins.

Ras is a signal transducting molecule involved in the receptor tyrosine kinase/RAS/Map kinase signalling cascade. Ras proteins bind GDP/GTP and show intrinsic GTPase activity. Mutations in ras, which change aa 12, 13 or 61 activate the potential of ras to transform cultured cells and are implicated in a variety of human tumours. The novel protein seems to be a new ras inhibitor protein.

The new protein can find application in modulating/blocking ras dependent signal transduction pathways.

Ras inhibitor

additional 1188 Bp at 5' and 1107 at 3' end in comparison to I22483

Sequenced by AGOWA

Locus: unknown

Insert length: 4137 bp
Poly A stretch at pos. 4116, no polyadenylation signal found

```
   1 GGGAGAACTG AAACAGGAGA TGGTGCGGAC AGATGTCAAC CTGGAAAATG
  51 GCCTGGAACC CGCTGAAACC CACAGCATGG TAAGACACAA GGATGGTGGC
 101 TATTCCGAGG AAGAGGACGT GAAGACCTGT GCCCGGGACT CAGGCTATGA
 151 CAGCCTCTCC AACAGGCTCA GCATCTTGGA CCGGCTCCTC CACACCCACC
 201 CCATATGGCT GCAGCTGAGT CTGAGTGAGG AGGAGGCAGC AGAGGTCCTG
 251 CAGGCCCAGC CTCCGGGGAT CTTCCTGGTT CATAAATCTA CCAAGATGCA
 301 GAAGAAAGTC CTCTCCCTCC GCCTGCCCTG TGAATTTGGG GCCCCACTCA
 351 AGGAATTTGC CATAAAGGAA AGCACATACA CCTTTTCCCT GGAAGGCTCA
 401 GGAATCAGTT TCGCAGATTT ATTCCGGCTC ATTGCTTTCT ACTGCATCAG
 451 CAGGGATGTT CTACCATTTA CCTTGAAGTT GCCTTATGCC ATTTCAACAG
 501 CCAAGTCGGA GGCTCAGCTT GAAGAACTGG CCCAGATGGG ACTAAATTTC
 551 TGGAGCTCCC CAGCTGACAG CAAACCCCCG AACCTTCCAC CTCCCCATAG
 601 GCCTCTTTCC TCCGACGGTG TCTGTCCTGC CTCCCTGCGT CAGCTCTGCC
 651 TTATAAATGG AGTGCATTCT ATCAAAACCA GGACGCCTTC AGAGCTGGAG
 701 TGCAGCCAGA CCAACGGGGC CCTGTGCTTT ATTAATCCCC TTTTCTTGAA
 751 AGTGCACAGC CAGGACCTCA GTGGAGGCCT GAAACGGCCG AGCACAAGGA
 801 CTCCCAACGC GAATGGCACG GAGCGGACTC GGTCCCCCCC ACCCAGGCCC
 851 CCGCCACCCG CTATTAATAG TCTCCACACA AGCCCTCGGC TGGCCAGGAC
 901 TGAAACCCAG ACGAGCATGC CAGAAACAGT CAACCATAAC AAACATGGGA
 951 ACGTAGCTCT GCCTGGAACG AAACCAACTC CCATCCCTCC ACCCCGGCTG
1001 AAGAAGCAGG CTTCTTTTCT GGAAGCAGAG GGCGGTGCAA AGACCTTGAG
1051 CGGCGGCCGG CCGGGCGCAG GCCCGGAGCT GGAGCTGGGC ACAGCTGGCA
1101 GCCCAGGTGG GGCCCCGCCT GAGGCCGCCC CGGGGGATTG CACAAGGGCC
1151 CCGCCGCCCA GCTCTGAATC ACGGCCCCCG TGCCATGGAG GCCGGCAGCG
1201 GCTGAGCGAC ATGAGCATTT CTACTTCCTC CTCCGACTCG CTGGAGTTCG
1251 ACCGGAGCAT GCCTCTGTTT GGCTACGAGG CGGACACCAA CAGCAGCCTG
1301 GAGGACTACG AGGGGGAAAG TGACCAAGAG ACCATGGCGC CCCCCATCAA
1351 GTCCAAAAAG AAAAGGAGCA GCTCCTTCGT GCTGCCCAAG CTCGTCAAGT
1401 CCCAGCTGCA GAAGGTGAGC GGGGTGTTCA GCTCCTTCAT GACCCCGGAG
1451 AAGCGGATGG TCCGCAGGAT CGCCGAGCTT TCCCGGGACA AATGCACCTA
1501 CTTCGGGTGC TTAGTGCAGG ACTACGTGAG CTTCCTGCAG GAGAACAAGG
1551 AGTGCCACGT GTCCAGCACC GACATGCTGC AGACCATCCG GCAGTTCATG
1601 ACCCAGGTCA AGAACTATTT GTCTCAGAGC TCGGAGCTGG ACCCCCCCAT
1651 CGAGTCGCTG ATCCCTGAAG ACCAAATAGA TGTGGTGCTG GAAAAAGCCA
1701 TGCACAAGTG CATCTTGAAG CCCCTCAAGG GGCATGTGGA GGCCATGCTG
1751 AAGGACTTTC ACATGGCCGA TGGCTCATGG AAGCAACTCA AGGAGAACCT
1801 GCAGCTTGTG CGGCAGAGGA ATCCGCAGGA GCTGGGGGTC TTCGCCCCGA
1851 CCCCTGATTT TGTGGATGTG GAGAAAATCA AAGTCAAGTT CATGACCATG
1901 CAGAAGATGT ATTCGCCGGA AAAGAAGGTC ATGCTGCTGC TGCGGGTCTG
1951 CAAGCTCATT TACACGGTCA TGGAGAACAA CTCAGGGAGG ATGTATGGCG
2001 CTGATGACTT CTTGCCAGTC CTGACCTATG TCATAGCCCA GTGTGACATG
2051 CTTGAATTGG ACACTGAAAT CGAGTACATG ATGGAGCTCC TAGACCCATC
2101 GCTGTTACAT GGAGAAGGAG GCTATTACTT GACAAGCGCA TATGGAGCAC
2151 TTTCTCTGAT AAAGAATTTC CAAGAAGAAC AAGCAGCGCG ACTGCTCAGC
2201 TCAGAAACCA GAGACACCCT GAGGCAGTGG CACAAACGGA GAACCACCAA
2251 CCGGACCATC CCCTCTGTGG ACGACTTCCA GAATTACCTC CGAGTTGCAT
2301 TTCAGGAGGT CAACAGTGGT TGCACAGGAA AGACCCTCCT TGTGAGACCT
2351 TACATCACCA CTGAGGATGT GTGTCAGATC TGCGCTGAGA AGTTCAAGGT
2401 GGGGGACCCT GAGGAGTACA GCCTCTTTCT CTTCGTTGAC GAGACATGGC
2451 AGCAGCTGGC AGAGGACACT TACCCTCAAA AAATCAAGGC GGAGCTGCAC
```

```
2501 AGCCGACCAC AGCCCCACAT CTTCCACTTT GTCTACAAAC GCATCAAGAA
2551 CGATCCTTAT GGCATCATTT TCCAGAACGG GGAAGAAGAC CTCACCACCT
2601 CCTAGAAGAC AGGCGGGACT TCCCAGTGGT GCATCCAAAG GGGAGCTGGA
2651 AGCCTTGCCT TCCCGCTTCT ACATGCTTGA GCTTGAAAAG CAGTCACCTC
2701 CTCGGGGACC CCTCAGTGTA GTGACTAAGC CATCCACAGG CCAACTCGGC
2751 CAAGGGCAAC TTTAGCCACG CAAGGTAGCT GAGGTTTGTG AAACAGTAGG
2801 ATTCTCTTTT GGCAATGGAG AATTGCATCT GATGGTTCAA GTGTCCTGAG
2851 ATTGTTTGCT ACCTACCCCC AGTCAGGTTC TAGGTTGGCT TACAGGTATG
2901 TATATGTGCA GAAGAAACAC TTAAGATACA AGTTCTTTTG AATTCAACAG
2951 CAGATGCTTG CGATGCAGTG CGTCAGGTGA TTCTCACTCC TGTGGATGGC
3001 TTCATCCCTG CCTTCCTTCC TTTCTTTTTC CTTTTTTTTT TTTTTTTTTT
3051 TTTTTACAAA GAGCCTTCAT GTTTTTATAT ATTTCATAGA AATTTTTATA
3101 GCAGTTGCAG GTAAACTGTC AGGATTGGTT TTAAAATATT TTTGTAACTT
3151 TAAAATATTC TATAATTATG CATGTGATTT TAACATTTAA TATTCAAAAA
3201 TAAATCTCTT GCTGGATTTG AGAGTATTGC ATTTTTAAAG TCTCTCTTCT
3251 GTAACTGGAT GTTTTGGCAA CTTTGTGGGG AGAGACTGCT GGATTTCTTA
3301 AAGCAACGTA TTCCTGACAC TGGCCACAGA ATGCCTTTGG AAATCGGATG
3351 TACTGTTCTC TTGTTCACGT TTAGTGGTGT TTTGCTGTTT TGTTTTTTAA
3401 ACAAATGATG CTGAGAATAA GGAGAGAAAT GAATGTAGAG AGAGGTAGAG
3451 AGAGAAATAT GAACTCTAAC AAAGGACTGA GGAGTGCAGT CTGCTGGTTC
3501 AGGCTCTTCA AAAGATGTAG AAAAAGAGAT AGAAGGAACC ACCTATGCTT
3551 AAAATACTGT AAATATGCAG TGAGGTTTGG CAAAATCTAT TCCATGTGTG
3601 ATTTGCTTGT AGAAACAATT TTGAAAGCCC CTTGAGGAAA ATAAAAATCA
3651 AGAAGAACAC TTTTCTCCCT TTTCCATACA AATTAAAACT TAACAGCATC
3701 AAATTATTGG GACCAGAAAC CAAGTAATGT ATAATGTGGC TTTTGTTGAG
3751 TTAAATAAGA TGCTATATAA TGGAGAAGAA TTTGAAAATG CACAAAAAAA
3801 TCAATCTACA TTATCAGAAC CTGCAGTGAA ATTAAACTTA TGTTAAATAA
3851 AACCAGTTTG CAGGTGCACA AACTATGAGG GTCTTGTATC CACGTAACAC
3901 AGGTAGTTAC AAAAACATGT TATTGTACTG TGTAAAGATG CATAGTCATC
3951 TCATTTGGTT GGCTTTGTAC CTTGTACCTT TTTTAGCCTT GGCTTTTGTT
4001 GAACTAGAAC CCTCAGCACA TACTGTGTTG TACTTTTGTA AATGATTTTT
4051 TAAATGGAAT TTTGCACATA ATACATTGTA ATACTGTATG ATAATCATGT
4101 GTGAAAATAA TTTTTGAAAT AAAAAAAAAA AAAAAAA
```

BLAST Results

Entry I22483 from database EMBL:
Sequence 15 from patent US 5527896.
Length = 1829
Plus Strand HSPs:
Score = 9097 (1364.9 bits), Expect = 0.0, P = 0.0
Identities = 1821/1823 (99%), Positives = 1821/1823 (99%),

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2602 bp; peptide length: 861
Category: known protein
Classification: Cell signaling/communication

```
  1 MVRTDVNLEN GLEPAETHSM VRHKDGGYSE EEDVKTCARD SGYDSLSNRL
 51 SILDRLLHTH PIWLQLSLSE EEAAEVLQAQ PPGIFLVHKS TKMQKKVLSL
101 RLPCEFGAPL KEFAIKESTY TFSLEGSGIS FADLFRLIAF YCISRDVLPF
151 TLKLPYAIST AKSEAQLEEL AQMGLNFWSS PADSKPPNLP PPHRPLSSDG
201 VCPASLRQLC LINGVHSIKT RTPSELECSQ TNGALCFINP LFLKVHSQDL
251 SGGLKRPSTR TPNANGTERT RSPPPRPPPP AINSLHTSPR LARTETQTSM
301 PETVNHNKHG NVALPGTKPT PIPPPRLKKQ ASFLEAEGGA KTLSGGRPGA
351 GPELELGTAG SPGGAPPEAA PGDCTRAPPP SSESRPPCHG GRQRLSDMSI
401 STSSSDSLEF DRSMPLFGYE ADTNSSLEDY EGESDQETMA PPIKSKKKRS
451 SSFVLPKLVK SQLQKVSGVF SSFMTPEKRM VRRIAELSRD KCTYFGCLVQ
501 DYVSFLQENK ECHVSSTDML QTIRQFMTQV KNYLSQSSEL DPPIESLIPE
551 DQIDVVLEKA MHKCILKPLK GHVEAMLKDF HMADGSWKQL KENLQLVRQR
601 NPQELGVFAP TPDFVDVEKI KVKFMTMQKM YSPEKKVMLL LRVCKLIYTV
651 MENNSGRMYG ADDFLPVLTY VIAQCDMLEL DTEIEYMMEL LDPSLLHGEG
701 GYYLTSAYGA LSLIKNFQEE QAARLLSSET RDTLRQWHKR RTTNRTIPSV
751 DDFQNYLRVA FQEVNSGCTG KTLLVRPYIT TEDVCQICAE KFKVGDPEEY
801 SLFLFVDETW QQLAEDTYPQ KIKAELHSRP QPHIFHFVYK RIKNDPYGII
851 FQNGEEDLTT S
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_20g21, frame 2

TREMBL:RNU80076_1 product: "RIN1"; Rattus norvegicus RIN1 mRNA, complete cds., N = 3, Score = 606, P = 6.8e-97

PIR:A38637 Ras interactor RIN1 - human, N = 3, Score = 587, P = 1.9e-92

TREMBL:HSRASINL_1 product: "ras inhibitor"; Human ras inhibitor mRNA, 3' end., N = 2, Score = 592, P = 9.8e-61

SWISSPROT:RIN1_HUMAN RAS INTERACTION/INTERFERENCE PROTEIN 1 (RAS INHIBITOR JC99) (FRAGMENT)., N = 2, Score = 587, P = 4.1e-60

PIR:B38637 Ras inhibitor (clone JC265) - human (fragment), N = 1, Score = 2446, P = 4.6e-254

>PIR:B38637 Ras inhibitor (clone JC265) - human (fragment)
Length = 471

HSPs:

Score = 2446 (367.0 bits), Expect = 4.6e-254, P = 4.6e-254
Identities = 471/471 (100%), Positives = 471/471 (100%)

```
Query:  391 GRQRLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETMAPPIKSKKKRS 450
            GRQRLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETMAPPIKSKKKRS
Sbjct:    1 GRQRLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETMAPPIKSKKKRS 60

Query:  451 SSFVLPKLVKSQLQKVSGVFSSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK 510
            SSFVLPKLVKSQLQKVSGVFSSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK
Sbjct:   61 SSFVLPKLVKSQLQKVSGVFSSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK 120

Query:  511 ECHVSSTDMLQTIRQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLEKAMHKCILKPLK 570
            ECHVSSTDMLQTIRQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLEKAMHKCILKPLK
Sbjct:  121 ECHVSSTDMLQTIRQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLEKAMHKCILKPLK 180

Query:  571 GHVEAMLKDFHMADGSWKQLKENLQLVRQRNPQELGVFAPTPDFVDVEKIKVKFMTMQKM 630
            GHVEAMLKDFHMADGSWKQLKENLQLVRQRNPQELGVFAPTPDFVDVEKIKVKFMTMQKM
Sbjct:  181 GHVEAMLKDFHMADGSWKQLKENLQLVRQRNPQELGVFAPTPDFVDVEKIKVKFMTMQKM 240

Query:  631 YSPEKKVMLLLRVCKLIYTVMENNSGRMYGADDFLPVLTYVIAQCDMLELDTEIEYMMEL 690
            YSPEKKVMLLLRVCKLIYTVMENNSGRMYGADDFLPVLTYVIAQCDMLELDTEIEYMMEL
Sbjct:  241 YSPEKKVMLLLRVCKLIYTVMENNSGRMYGADDFLPVLTYVIAQCDMLELDTEIEYMMEL 300

Query:  691 LDPSLLHGEGGYYLTSAYGALSLIKNFQEEQAARLLSSETRDTLRQWHKRRTTNRTIPSV 750
            LDPSLLHGEGGYYLTSAYGALSLIKNFQEEQAARLLSSETRDTLRQWHKRRTTNRTIPSV
Sbjct:  301 LDPSLLHGEGGYYLTSAYGALSLIKNFQEEQAARLLSSETRDTLRQWHKRRTTNRTIPSV 360

Query:  751 DDFQNYLRVAFQEVNSGCTGKTLLVRPYITTEDVCQICAEKFKVGDPEEYSLFLFVDETW 810
            DDFQNYLRVAFQEVNSGCTGKTLLVRPYITTEDVCQICAEKFKVGDPEEYSLFLFVDETW
Sbjct:  361 DDFQNYLRVAFQEVNSGCTGKTLLVRPYITTEDVCQICAEKFKVGDPEEYSLFLFVDETW 420

Query:  811 QQLAEDTYPQKIKAELHSRPQPHIFHFVYKRIKNDPYGIIFQNGEEDLTTS 861
            QQLAEDTYPQKIKAELHSRPQPHIFHFVYKRIKNDPYGIIFQNGEEDLTTS
Sbjct:  421 QQLAEDTYPQKIKAELHSRPQPHIFHFVYKRIKNDPYGIIFQNGEEDLTTS 471
```

Pedant information for DKFZphutel_20g21, frame 2

Report for DKFZphutel_20g21.2

[LENGTH]	861
[MW]	96380.26
[pI]	6.15
[HOMOL]	PIR:B38637 Ras inhibitor (clone JC265) - human (fragment) 0.0
[FUNCAT]	08.13 vacuolar transport [S. cerevisiae, YML097c] 3e-10
[FUNCAT]	06.04 protein targeting, sorting and translocation [S. cerevisiae, YML097c] 3e-10
[FUNCAT]	30.03 organization of cytoplasm [S. cerevisiae, YML097c] 3e-10
[FUNCAT]	08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YML097c] 3e-10
[PIRKW]	alternative splicing 3e-59
[SUPFAM]	Ras interactor RIN1 3e-59

```
[KW]        All_Alpha
[KW]        LOW_COMPLEXITY     11.27 %

SEQ    MVRTDVNLENGLEPAETHSMVRHKDGGYSEEEDVKTCARDSGYDSLSNRLSILDRLLHTH
SEG    ............................................................
PRD    cccccceeeccccccccccceeeeeecccccccccccceeeeeecccccchhhhhhhhhhhhhh

SEQ    PIWLQLSLSEEEAAEVLQAQPPGIFLVHKSTKMQKKVLSLRLPCEFGAPLKEFAIKESTY
SEG    ...xxxxxxxxxxxxxxxxx........................................
PRD    hhhhhhhhhhhhhhhhhhhhhcccccceeeeechhhhhhhhhhhcccccccccceeeeeeecc

SEQ    TFSLEGSGISFADLFRLIAFYCISRDVLPFTLKLPYAISTAKSEAQLEELAQMGLNFWSS
SEG    ............................................................
PRD    ceeeccccchhhhhhhhhhhhhcceeeeeeecccchhhhhhhhhhhhhhhhhhhcccccccc

SEQ    PADSKPPNLPPPHRPLSSDGVCPASLRQLCLINGVHSIKTRTPSELECSQTNGALCFINP
SEG    .....xxxxxxxxxx.............................................
PRD    ccccccccccccccccccccccccchhhhhhccccccccccccccccccccccccceeeecc

SEQ    LFLKVHSQDLSGGLKRPSTRTPNANGTERTRSPPPRPPPPAINSLHTSPRLARTETQTSM
SEG    ................................xxxxxxxx....................
PRD    cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ    PETVNHNKHGNVALPGTKPTPIPPPRLKKQASFLEAEGGAKTLSGGRPGAGPELELGTAG
SEG    ..............xxxxxxxxxxx...................................xx
PRD    eeeeeccccccccccccccccccccccchhhhhhhhhhhcccccccccccccccccceeeeeccc

SEQ    SPGGAPPEAAPGDCTRAPPPSSESRPPCHGGRQRLSDMSISTSSSDSLEFDRSMPLFGYE
SEG    xxxxxxxxxxxx.....xxxxxxxxxx.......xxxxxxxxxxxxxxxxx.......
PRD    ccccccccccccccccccccccccccccccccccccccccccccccccceeecccccccceee

SEQ    ADTNSSLEDYEGESDQETMAPPIKSKKKRSSSFVLPKLVKSQLQKVSGVFSSFMTPEKRM
SEG    ......................xxxxxxxxx.............................
PRD    cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhcchhhh

SEQ    VRRIAELSRDKCTYFGCLVQDYVSFLQENKECHVSSTDMLQTIRQFMTQVKNYLSQSSEL
SEG    ............................................................
PRD    hhhhhhhhhhchhhhhhhhhhhhhhhhhhcccccccccchhhhhhhhhhhhhhhhhhhhhcc

SEQ    DPPIESLIPEDQIDVVLEKAMHKCILKPLKGHVEAMLKDFHMADGSWKQLKENLQLVRQR
SEG    ............................................................
PRD    ccccccccccchhhhhhhhhhhhcccccchhhhhhhhhhhhhccccchhhhhhhhhhhhh

SEQ    NPQELGVFAPTPDFVDVEKIKVKFMTMQKMYSPEKKVMLLLRVCKLIYTVMENNSGRMYG
SEG    ............................................................
PRD    cccccccccccccchhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhcccccc

SEQ    ADDFLPVLTYVIAQCDMLELDTEIEYMMELLDPSLLHGEGGYYLTSAYGALSLIKNFQEE
SEG    ............................................................
PRD    ccccccccceeecccccchhhhhhhhhhhhhhcccccccccceeeeehhhhhhhhhhhhh

SEQ    QAARLLSSETRDTLRQWHKRRTTNRTIPSVDDFQNYLRVAFQEVNSGCTGKTLLVRPYIT
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhhhhcccccccchhhhhhhhhhccccccceeeeeeccccc

SEQ    TEDVCQICAEKFKVGDPEEYSLFLFVDETWQQLAEDTYPQKIKAELHSRPQPHIFHFVYK
SEG    ............................................................
PRD    chhhhhhhhheeeccccccceeeeehhhhhhccccccchhhhhhhhhccccceeeehhh

SEQ    RIKNDPYGIIFQNGEEDLTTS
SEG    .....................
PRD    hhcccccceeeeecccccccc
```

(No Prosite data available for DKFZphutel_20g21.2)

(No Pfam data available for DKFZphutel_20g21.2)

DKFZphutel_20h13
----

group: intracellular transport and trafficking

DKFZphutel_20h13 encodes a novel 955 amino acid protein with similarity to alpha-adaptins.

Adaptins are components of the adaptor complexes which link clathrin to receptors in coated vesicles. The alpha-adaptins, which are found exclusively in endocytic coated vesicles, separate into two bands on SDS gels, designated A and C. The novel protein is very similar to both alpha adaptin A and C. The novel protein is a new human alpha-adaptin.

The new protein can find application in modulating endocytosis and vesicle trafficking in cells.

strong similarity to alpha-adaptins

complete cDNA, complete cds start at Bp 78, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 3352 bp
Poly A stretch at pos. 3297, polyadenylation signal at pos. 3279

```
   1 GCGCCCGGTC CCCGCTTGCC AGCCCCCGCT GCTCTGTGCC CTGTCCGGCC
  51 AGGCCTGGAG CCGACACCAC CGCCATCATG CCGGCCGTGT CCAAGGGCGA
 101 TGGGATGCGG GGGCTCGCGG TGTTCATCTC CGACATCCGG AACTGTAAGA
 151 GCAAAGAGGC GGAAATTAAG AGAATCAACA AGGAACTGGC CAACATCCGC
 201 TCCAAGTTCA AAGGAGACAA AGCCTTGGAT GGCTACAGTA AGAAAAAATA
 251 TGTGTGTAAA CTGCTTTTCA TCTTCCTGCT TGGCCATGAC ATTGACTTTG
 301 GGCACATGGA GGCTGTGAAT CTGTTGAGTT CCAATAAATA CACAGAGAAG
 351 CAAATAGGTT ACCTGTTCAT TTCTGTGCTG GTGAACTCGA ACTCGGAGCT
 401 GATCCGCCTC ATCAACAACG CCATCAAGAA TGACCTGGCC AGCCGCAACC
 451 CCACCTTCAT GTGCCTGGCC CTGCACTGCA TCGCCAACGT GGGCAGCCGG
 501 GAGATGGGCG AGGCCTTTGC CGCTGACATC CCCCGCATCC TGGTGGCCGG
 551 GGACAGCATG GACAGTGTCA AGCAGAGTGC GGCCCTGTGC CTCCTTCGAC
 601 TGTACAAGGC CTCGCCTGAC CTGGTGCCCA TGGGCGAGTG GACGGCGCGT
 651 GTGGTACACC TGCTCAATGA CCAGCACATG GGTGTGGTCA CGGCCGCCGT
 701 CAGCCTCATC ACCTGTCTCT GCAAGAAGAA CCCAGATGAC TTCAAGACGT
 751 GCGTCTCTCT GGCTGTGTCG CGCCTGAGCC GGATCGTCTC CTCTGCCTCC
 801 ACCGACCTCC AGGACTACAC CTACTACTTC GTCCCAGCAC CCTGGCTCTC
 851 GGTGAAGCTC CTGCGGCTGC TGCAGTGCTA CCCGCCTCCA GAGGATGCGG
 901 CTGTGAAGGG GCGGCTGGTG GAATGTCTGG AGACTGTGCT CAACAAGGCC
 951 CAGGAGCCCC CCAAATCCAA GAAGGTGCAG CATTCCAACG CCAAGAACGC
1001 CATCCTCTTC GAGACCATCA GCCTCATCAT CCACTATGAC AGTGAGCCCA
1051 ACCTCCTGGT TCGGGCCTGC AACCAGCTGG GCCAGTTCCT GCAGCACCGG
1101 GAGACCAACC TGCGCTACCT GGCCCTGGAG AGCATGTGCA CGCTGGCCAG
1151 CTCCGAGTTC TCCCATGAAG CCGTCAAGAC GCACATTGAC ACCGTCATCA
1201 ATGCCCTCAA GACGGAGCGG GACGTCAGCG TGCGGCAGCG GGCGGCTGAC
1251 CTCCTCTACG CCATGTGTGA CCGGAGCAAT GCCAAGCAGA TCGTGTCGGA
1301 GATGCTGCGG TACCTGGAGA CGGCAGACTA CGCCATCCGC GAGGAGATCG
1351 TCCTGAAGGT GGCCATCCTG GCCGAGAAGT ACGCCGTGGA CTACAGCTGG
1401 TACGTGGACA CCATCCTCAA CCTCATCCGC ATTGCGGGCG ACTACGTGAG
1451 TGAGGAGGTG TGGTACCGTG TGCTACAGAT CGTCACCAAC CGTGATGACG
1501 TCCAGGGCTA TGCCGCCAAG ACCGTCTTTG AGGCGCTCCA GGCCCCTGCC
1551 TGTCACGAGA ACATGGTGAA GGTTGGCGGC TACATCCTTG GGGAGTTTGG
1601 GAACCTGATT GCTGGGGACC CCCGCTCCAG CCCCCCAGTG CAGTTCTCCC
1651 TGCTCCACTC CAAGTTCCAT CTGTGCAGCG TGGCCACGCG GGCGCTGCTG
1701 CTGTCCACCT ACATCAAGTT CATCAACCTC TTCCCCGAGA CCAAGGCCAC
1751 CATCCAGGGC GTCCTGCGGG CCGGCTCCCA GCTGCGCAAT GCTGACGTGG
1801 AGCTGCAGCA GCGAGCCGTG GAGTACCTCA CCCTCAGCTC AGTGGCCAGC
1851 ACCGACGTCC TGGCCACGGT GCTGGAGGAG ATGCCGCCCT TCCCCGAGCG
1901 CGAGTCGTCC ATCCTGGCCA AGCTGAAACG CAAGAAGGGG CCAGGGGCCG
1951 GCAGCGCCCT GGACGATGGC CGGAGGGACC CCAGCAGCAA CGACATCAAC
2001 GGGGGCATGG AGCCCACCCC CAGCACTGTG TCGACGCCCT CGCCCTCCGC
2051 CGACCTCCTG GGGCTGCGGG CAGCCCCTCC CCCGGCAGCA CCCCCGGCTT
2101 CTGCAGGAGC AGGGAACCTT CTGGTGGACG TCTTCGATGG CCCGGCCGCC
2151 CAGCCCAGCC TGGGGCCCAC CCCCGAGGAG GCCTTCCTCA GCCCAGGTCC
2201 TGAGGACATC GGCCCTCCCA TTCCGGAAGC CGATGAGTTG CTGAATAAGT
2251 TTGTGTGTAA GAACAACGGG GTCCTGTTCG AGAACCAGCT GCTGCAGATC
2301 GGAGTCAAGT CAGAGTTCCG ACAGAACCTG GGCCGCATGT ATCTCTTCTA
2351 TGGCAACAAG ACCTCGGTGC AGTTCCAGAA TTTCTCACCC ACTGTGGTTC
2401 ACCCGGGAGA CCTCCAGACT CAGCTGGCTG TGCAGACCAA GCGCGTGGCG
2451 GCGCAGGTGG ACGGCGGCGC GCAGGTGCAG CAGGTGCTCA ATATCGAGTG
2501 CCTGCGGGAC TTCCTGACGC CCCCGCTGCT GTCCGTGCGC TTCCGGTACG
2551 GTGGCGCCCC CCAGGCCCTC ACCCTGAAGC TCCCAGTGAC CATCAACAAG
```

```
2601 TTCTTCCAGC CCACCGAGAT GGCGGCCCAG GATTTCTTCC AGCGCTGGAA
2651 GCAGCTGAGC CTCCCTCAAC AGGAGGCGCA GAAAATCTTC AAAGCCAACC
2701 ACCCCATGGA CGCAGAAGTT ACTAAGGCCA AGCTTCTGGG GTTTGGCTCT
2751 GCTCTCCTGG ACAATGTGGA CCCCAACCCT GAGAACTTCG TGGGGGCGGG
2801 GATCATCCAG ACTAAAGCCC TGCAGGTGGG CTGTCTGCTT CGGCTGGAGC
2851 CCAATGCCCA GGCCCAGATG TACCGGCTGA CCCTGCGCAC CAGCAAGGAG
2901 CCCGTCTCCC GTCACCTGTG TGAGCTGCTG GCACAGCAGT TCTGAGCCCT
2951 GGACTCTGCC CCGGGGGATG TGGCCGGCAC TGGGCAGCCC CTTGGACTGA
3001 GGCAGTTTTG GTGGATGGGG GACCTCCACT GGTGACAGAG AAGACACCAG
3051 GGTTTGGGGG ATGCCTGGGA CTTTCCTCCG GCCTTTTGTA TTTTTATTTT
3101 TGTTCATCTG CTGCTGTTTA CATTCTGGGG GGTTAGGGGG AGTCCCCCTC
3151 CCTCCCTTTC CCCCCCAAGC ACAGAGGGGA GAGGGGCCAG GGAAGTGGAT
3201 GTCTCCTCCC CTCCCACCCC ACCCTGTTGT AGCCCCTCCT ACCCCCTCCC
3251 CATCCAGGGG CTGTGTATTA TTGTGAGCGA ATAAACAGAG AGACGCTAAA
3301 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AA
```

BLAST Results

No BLAST result

Medline entries

89155572:
Cloning of cDNAs encoding two related 100-kD coated vesicle proteins (alpha-adaptins).

97431776:
Alpha-adaptin, a marker for endocytosis, is expressed in complex patterns during Drosophila development.

Peptide information for frame 3

ORF from 78 bp to 2942 bp; peptide length: 955
Category: strong similarity to known protein

```
  1 MPAVSKGDGM RGLAVFISDI RNCKSKEAEI KRINKELANI RSKFKGDKAL
 51 DGYSKKKYVC KLLFIFLLGH DIDFGHMEAV NLLSSNKYTE KQIGYLFISV
101 LVNSNSELIR LINNAIKNDL ASRNPTFMCL ALHCIANVGS REMGEAFAAD
151 IPRILVAGDS MDSVKQSAAL CLLRLYKASP DLVPMGEWTA RVVHLLNDQH
201 MGVVTAAVSL ITCLCKKNPD DFKTCVSLAV SRLSRIVSSA STDLQDYTYY
251 FVPAPWLSVK LLRLLQCYPP PEDAAVKGRL VECLETVLNK AQEPPKSKKV
301 QHSNAKNAIL FETISLIIHY DSEPNLLVRA CNQLGQFLQH RETNLRYLAL
351 ESMCTLASSE FSHEAVKTHI DTVINALKTE RDVSVRQRAA DLLYAMCDRS
401 NAKQIVSEML RYLETADYAI REEIVLKVAI LAEKYAVDYS WYVDTILNLI
451 RIAGDYVSEE VWYRVLQIVT NRDDVQGYAA KTVFEALQAP ACHENMVKVG
501 GYILGEFGNL IAGDPRSSPP VQFSLLHSKF HLCSVATRAL LLSTYIKFIN
551 LFPETKATIQ GVLRAGSQLR NADVELQQRA VEYLTLSSVA STDVLATVLE
601 EMPPFPERES SILAKLKRKK GPGAGSALDD GRRDPSSNDI NGGMEPTPST
651 VSTPSPSADL LGLRAAPPPA APPASAGAGN LLVDVFDGPA AQPSLGPTPE
701 EAFLSPGPED IGPPIPEADE LLNKFVCKNN GVLFENQLLQ IGVKSEFRQN
751 LGRMYLFYGN KTSVQFQNFS PTVVHPGDLQ TQLAVQTKRV AAQVDGGAQV
801 QQVLNIECLR DFLTPPLLSV RFRYGGAPQA LTLKLPVTIN KFFQPTEMAA
851 QDFFQRWKQL SLPQQEAQKI FKANHPMDAE VTKAKLLGFG SALLDNVDPN
901 PENFVGAGII QTKALQVGCL LRLEPNAQAQ MYRLTLRTSK EPVSRHLCEL
951 LAQQF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_20h13, frame 3

PIR:B30111 alpha-adaptin C - mouse, N = 1, Score = 3990, P = 0

PIR:S11276 alpha-adaptin c - rat, N = 1, Score = 3987, P = 0

SWISSPROT:ADAC_RAT ALPHA-ADAPTIN C (CLATHRIN ASSEMBLY PROTEIN COMPLEX 2 ALPHA-C LARGE CHAIN) (100 KD COATED VESICLE PROTEIN C) (PLASMA MEMBRANE ADAPTOR HA2/AP2 ADAPTIN ALPHA C SUBUNIT)., N = 1, Score = 3982, P = 0

SWISSPROT:ADAC_MOUSE ALPHA-ADAPTIN C (CLATHRIN ASSEMBLY PROTEIN COMPLEX 2 ALPHA-C LARGE CHAIN) (100 KD COATED VESICLE PROTEIN C) (PLASMA MEMBRANE ADAPTOR HA2/AP2 ADAPTIN ALPHA C SUBUNIT)., N = 1, Score = 3976, P = 0

TREMBL:AB020706_1 gene: "KIAA0899"; product: "KIAA0899 protein"; Homo sapiens mRNA for KIAA0899 protein, partial cds., N = 1, Score = 3932, P = 0

```
>PIR:B30111 alpha-adaptin C - mouse
          Length = 938

  HSPs:

 Score = 3990 (598.6 bits), Expect = 0.0e+00, P = 0.0e+00
 Identities = 787/955 (82%), Positives = 858/955 (89%)

Query:     1 MPAVSKGDGMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC 60
             MPAVSKGDGMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC
Sbjct:     1 MPAVSKGDGMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC 60

Query:    61 KLLFIFLLGHDIDFGHMEAVNLLSSNKYTEKQIGYLFISVLVNSNSELIRLINNAIKNDL 120
             KLLFIFLLGHDIDFGHMEAVNLLSSN+YTEKQIGYLFISVLVNSNSELIRLINNAIKNDL
Sbjct:    61 KLLFIFLLGHDIDFGHMEAVNLLSSNRYTEKQIGYLFISVLVNSNSELIRLINNAIKNDL 120

Query:   121 ASRNPTFMCLALHCIANVGSREMGEAFAADIPRILVAGDSMDSVKQSAALCLLRLYKASP 180
             ASRNPTFM LALHCIANVGSREM EAFA +IP+ILVAGD+MDSVKQSAALCLLRLY+ SP
Sbjct:   121 ASRNPTFMGLALHCIANVGSREMAEAFAGEIPKILVAGDTMDSVKQSAALCLLRLYRTSP 180

Query:   181 DLVPMGEWTARVVHLLNDQHMGVVTAAVSLITCLCKKNPDDFKTCVSLAVSRLSRIVSSA 240
             DLVPMG+WT+RVVHLLNDQH+GVVTAA SLIT L +KNP++FKT VSLAVSRLSRIV+SA
Sbjct:   181 DLVPMGDWTSRVVHLLNDQHLGVVTAATSLITTLAQKNPEEFKTSVSLAVSRLSRIVTSA 240

Query:   241 STDLQDYTYYFVPAPWLSVKLLRLLQCYPPPEDAAVKGRLVECLETVLNKAQEPPKSKKV 300
             STDLQDYTYYFVPAPWLSVKLLRLLQCYPPP D AV+GRL ECLET+LNKAQEPPKSKKV
Sbjct:   241 STDLQDYTYYFVPAPWLSVKLLRLLQCYPPP-DPAVRGRLTECLETILNKAQEPPKSKKV 299

Query:   301 QHSNAKNAILFETISLIIHYDSEPNLLVRACNQLGQFLQHRETNLRYLALESMCTLASSE 360
             QHSNAKNA+LFE ISLIIH+DSEPNLLVRACNQLGQFLQHRETNLRYLALESMCTLASSE
Sbjct:   300 QHSNAKNAVLFEAISLIIHHDSEPNLLVRACNQLGQFLQHRETNLRYLALESMCTLASSE 359

Query:   361 FSHEAVKTHIDTVINALKTERDVSVRQRAADLLYAMCDRSNAKQIVSEMLRYLETADYAI 420
             FSHEAVKTHI+TVINALKTERDVSVRQRA DLLYAMCDRSNA+QIV+EML YLETADY+I
Sbjct:   360 FSHEAVKTHIETVINALKTERDVSVRQRAVDLLYAMCDRSNAQQIVAEMLSYLETADYSI 419

Query:   421 REEIVLKVAILAEKYAVDYSWYVDTILNLIRIAGDYVSEEVWYRVLQIVTNRDDVQGYAA 480
             REEIVLKVAILAEKYAVDY+WYVDTILNLIRIAGDYVSEEVWYRV+QIV NRDDVQGYAA
Sbjct:   420 REEIVLKVAILAEKYAVDYTWYVDTILNLIRIAGDYVSEEVWYRVIQIVINRDDVQGYAA 479

Query:   481 KTVFEALQAPACHENMVKVGGYILGEFGNLIAGDPRSSPPVQFSLLHSKFHLCSVATRAL 540
             KTVFEALQAPACHEN+VKVGGYILGEFGNLIAGDPRSSP +QF+LLHSKFHLCSV TRAL
Sbjct:   480 KTVFEALQAPACHENLVKVGGYILGEFGNLIAGDPRSSPLIQFNLLHSKFHLCSVPTRAL 539

Query:   541 LLSTYIKFINLFPETKATIQGVLRAGSQLRNADVELQQRAVEYLTLSSVASTDVLATVLE 600
             LLSTYIKF+NLFPE KATIQ VLR+ SQL+NADVELQQRAVEYL LS+VASTD+LATVLE
Sbjct:   540 LLSTYIKFVNLFPEVKATIQDVLRSDSQLKNADVELQQRAVEYLRLSTVASTDILATVLE 599

Query:   601 EMPPFPERESSILAKLKRKKGPGAGSALDDGRRDPSSNDINGGMEPTP---STVSTPSPS 657
             EMPPFPERESSILAKLK+KKGP    + L++ +R+ S  D+NGG EP P   S   STPSPS
Sbjct:   600 EMPPFPERESSILAKLKKKKGPSTVTDLEETKRERSI-DVNGGPEPVPASTSAASTPSPS 658

Query:   658 ADLLGLRAAPP-PAAPPASAGAGNLLVDVFDGPAAQPSLGPTPEEAFLSPGPEDIGPPIP 716
             ADLLGL A PP P  PP S+G G LLVDVF    A+  ++ P       L+PG ED
Sbjct:   659 ADLLGLGAVPPAPTGPPPSSGGG-LLVDVFSDSAS--AVAP------LAPGSEDN----- 704

Query:   717 EADELLNKFVCKNNGVLFENQLLQIGVKSEFRQNLGRMYLFYGNKTSVQFQNFSPTVVHP 776
                   +FVCKNNGVLFENQLLQIG+KSEFRQNLGRM+ FYGNKTS QF NF+PT++
Sbjct:   705 -----FARFVCKNNGVLFENQLLQIGLKSEFRQNLGRMFIFYGNKTSTQFLNFTPTLICA 759

Query:   777 GDLQTQLAVQTKRVAAQVDGGAQVQQVLNIECLRDFLTPPLLSVRFRYGGAPQALTLKLP 836
              DLQT L +QTK V   VDGGAQVQQV+NIEC+ DF   P+L+++FRYGG  Q +++KLP
Sbjct:   760 DDLQTNLNLQTKPVDPTVDGGAQVQQVVNIECISDFTEAPVLNIQFRYGGTFQNVSVKLP 819

Query:   837 VTINKFFQPTEMAAQDFFQRWKQLSLPQQEAQKIFKANHPMDAEVTKAKLLGFGSALLDN 896
             +T+NKFFQPTEMA+QDFFQRWKQLS PQQE Q IFKA HPMD E+TKAK++GFGSALL+
Sbjct:   820 ITLNKFFQPTEMASQDFFQRWKQLSNPQQEVQNIFKAKHPMDTEITKAKIIGFGSALLEE 879

Query:   897 VDPNPENFVGAGIIQTKALQVGCLLRLEPNAQAQMYRLTLRTSKEPVSRHLCELLAQQF 955
             VDPNP NFVGAGII TK  Q+GCLLRLEPN QAQMYRLTLRTSK+ VS+ LCELL++QF
```

Sbjct: 880 VDPNPANFVGAGIIHTKTTQIGCLLRLEPNLQAQMYRLTLRTSKDTVSQRLCELLSEQF 938

Pedant information for DKFZphutel_20h13, frame 3

Report for DKFZphutel_20h13.3

[LENGTH]	955
[MW]	105361.97
[pI]	7.75
[HOMOL]	PIR:A30111 alpha-adaptin A - mouse 0.0
[FUNCAT]	30.09 organization of intracellular transport vesicles [S. cerevisiae, YBL037w] 5e-67
[FUNCAT]	08.19 cellular import [S. cerevisiae, YBL037w] 5e-67
[FUNCAT]	06.10 assembly of protein complexes [S. cerevisiae, YBL037w] 5e-67
[FUNCAT]	08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDR238c] 4e-04
[PIRKW]	heterodimer 0.0
[PIRKW]	transmembrane protein 1e-65
[PIRKW]	membrane trafficking 0.0
[PIRKW]	receptor 0.0
[SUPFAM]	beta-adaptin 5e-16
[PROSITE]	MYRISTYL 7
[PROSITE]	IG_MHC 1
[PROSITE]	AMIDATION 1
[PROSITE]	CK2_PHOSPHO_SITE 11
[PROSITE]	TYR_PHOSPHO_SITE 3
[PROSITE]	PKC_PHOSPHO_SITE 15
[PROSITE]	ASN_GLYCOSYLATION 1
[KW]	All_Alpha
[KW]	LOW_COMPLEXITY 6.81 %

```
SEQ  MPAVSKGDGMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC
SEG  ............................................................
PRD  cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccchhhhhh

SEQ  KLLFIFLLGHDIDFGHMEAVNLLSSNKYTEKQIGYLFISVLVNSNSELIRLINNAIKNDL
SEG  ............................................................
PRD  hhhhhhhcccccccchhhhhhhhhcccccchhhhhhhhhhhhcchhhhhhhhhhhhhcc

SEQ  ASRNPTFMCLALHCIANVGSREMGEAFAADIPRILVAGDSMDSVKQSAALCLLRLYKASP
SEG  ............................................................
PRD  cccccchhhhhhhhhhcccchhhhhhhhhhhhhheeeccccchhhhhhhhhhhhhhhhcc

SEQ  DLVPMGEWTARVVHLLNDQHMGVVTAAVSLITCLCKKNPDDFKTCVSLAVSRLSRIVSSA
SEG  ............................................................
PRD  cccccccchhhhhhhhhcccceeeehhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhcc

SEQ  STDLQDYTYYFVPAPWLSVKLLRLLQCYPPPEDAAVKGRLVECLETVLNKAQEPPKSKKV
SEG  ............................................................
PRD  cccccccceeeecccchhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhcccccc

SEQ  QHSNAKNAILFETISLIIHYDSEPNLLVRACNQLGQFLQHRETNLRYLALESMCTLASSE
SEG  ............................................................
PRD  cccccchhhhhhhhhhhhcccccceeeeehhhhhhhhhcccccceeeehhhhhhhhcc

SEQ  FSHEAVKTHIDTVINALKTERDVSVRQRAADLLYAMCDRSNAKQIVSEMLRYLETADYAI
SEG  ............................................................
PRD  cchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhcccch

SEQ  REEIVLKVAILAEKYAVDYSWYVDTILNLIRIAGDYVSEEVWYRVLQIVTNRDDVQGYAA
SEG  ............................................................
PRD  hhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhcccccchhhhhhhheeeccccchhhhhh

SEQ  KTVFEALQAPACHENMVKVGGYILGEFGNLIAGDPRSSPPVQFSLLHSKFHLCSVATRAL
SEG  ............................................................
PRD  hhhhhhhhhhccccceeeeeeeeccccccccccccccchhhhhhhhhhhccccchhhhh

SEQ  LLSTYIKFINLFPETKATIQGVLRAGSQLRNADVELQQRAVEYLTLSSVASTDVLATVLE
SEG  ............................................................
PRD  hhhhhhhhhhcccccchhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhcccchhhhhhhhhh

SEQ  EMPPFPERESSILAKLKRKKGPGAGSALDDGRRDPSSNDINGGMEPTPSTVSTPSPSADL
SEG  .............................................xxxxxxxxxxxxxxx
PRD  hcccccccchhhhhhhhhccccccccccccccccccccccccccccccccccccccccce

SEQ  LGLRAAPPPAAPPASAGAGNLLVDVFDGPAAQPSLGPTPEEAFLSPGPEDIGPPIPEADE
SEG  xxxxxxxxxxxxxxxxxxxxx......................xxxxxxxxxxxxxx.
```

```
PRD     ecccccccccccccccccceeeeeeccccccccccccccccceeecccccccccccccccc

SEQ     LLNKFVCKNNGVLFENQLLQIGVKSEFRQNLGRMYLFYGNKTSVQFQNFSPTVVHPGDLQ
SEG     ............................................................
PRD     cceeeeeeccccccchhhhhhhcchhhhhccccceeeccccccccccccceeeeccchhh

SEQ     TQLAVQTKRVAAQVDGGAQVQQVLNIECLRDFLTPPLLSVRFRYGGAPQALTLKLPVTIN
SEG     .........xxxxxxxxxxxxxx.....................................
PRD     hhhhhhhcccccccccchhhhhhhhhcccccccceeeeeecccccccccccccccccccc

SEQ     KFFQPTEMAAQDFFQRWKQLSLPQQEAQKIFKANHPMDAEVTKAKLLGFGSALLDNVDPN
SEG     ............................................................
PRD     cccccchhhhhhhhhhhhhhchhhhhhhhhhhhcccchhhhhhhhhhccccceeeecccc

SEQ     PENFVGAGIIQTKALQVGCLLRLEPNAQAQMYRLTLRTSKEPVSRHLCELLAQQF
SEG     .......................................................
PRD     ccceeeceeeeeccccceeeeecccchhhhhhhhhhhccccchhhhhhhhhhccc
```

Prosite for DKFZphutel_20h13.3

PS00001	760->764	ASN_GLYCOSYLATION	PDOC00001
PS00005	54->57	PKC_PHOSPHO_SITE	PDOC00005
PS00005	85->88	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	163->166	PKC_PHOSPHO_SITE	PDOC00005
PS00005	189->192	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00005	297->300	PKC_PHOSPHO_SITE	PDOC00005
PS00005	379->382	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	470->473	PKC_PHOSPHO_SITE	PDOC00005
PS00005	787->790	PKC_PHOSPHO_SITE	PDOC00005
PS00005	819->822	PKC_PHOSPHO_SITE	PDOC00005
PS00005	832->835	PKC_PHOSPHO_SITE	PDOC00005
PS00005	935->938	PKC_PHOSPHO_SITE	PDOC00005
PS00005	938->941	PKC_PHOSPHO_SITE	PDOC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDOC00006
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	368->372	CK2_PHOSPHO_SITE	PDOC00006
PS00006	379->383	CK2_PHOSPHO_SITE	PDOC00006
PS00006	470->474	CK2_PHOSPHO_SITE	PDOC00006
PS00006	482->486	CK2_PHOSPHO_SITE	PDOC00006
PS00006	597->601	CK2_PHOSPHO_SITE	PDOC00006
PS00006	626->630	CK2_PHOSPHO_SITE	PDOC00006
PS00006	636->640	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00006	938->942	CK2_PHOSPHO_SITE	PDOC00006
PS00007	388->395	TYR_PHOSPHO_SITE	PDOC00007
PS00007	411->419	TYR_PHOSPHO_SITE	PDOC00007
PS00007	434->443	TYR_PHOSPHO_SITE	PDOC00007
PS00008	202->208	MYRISTYL	PDOC00008
PS00008	508->514	MYRISTYL	PDOC00008
PS00008	561->567	MYRISTYL	PDOC00008
PS00008	623->629	MYRISTYL	PDOC00008
PS00008	759->765	MYRISTYL	PDOC00008
PS00008	826->832	MYRISTYL	PDOC00008
PS00008	908->914	MYRISTYL	PDOC00008
PS00009	630->634	AMIDATION	PDOC00009
PS00290	127->134	IG_MHC	PDOC00262

(No Pfam data available for DKFZphutel_20h13.3)

DKFZphute1_20m11
----

group: cell cycle

DKFZphute1_20m11 encodes a novel 225 amino acid protein with similarity to yeast sds22 and protein phosphatase-1 regulatory subunits.

sds22 is a regulatory polypeptide of protein phosphatase-1 that is required for the completion of mitosis in both fission and budding yeast. The novel protein seems to be a new regulator protein for protein phosphatase-1.

The new protein can find application in modulating/blocking the activity of protein phosphatase-1 and in modulating the cell cycle.

similarity to suppressor protein sds22

complete cDNA, complete cds, EST hits
localisation? only a part of the STS matches

Sequenced by AGOWA

Locus: /map="17"?

Insert length: 5822 bp
Poly A stretch at pos. 5803, polyadenylation signal at pos. 5786

```
   1 GGGCGCTTGG TTCCCCAGCA ACCGGGAGAC GCGTCTGCTG CGTGGAACCG
  51 CCGAGTTCCC AGCGCTTGAG AAGGAAAATT CTGGATCTGT TATCTGTGAG
 101 GAGGCCACTC CGTTGACAGT TGTGTAAAAC TCTGCTGCTT TCCCCAGCTC
 151 CAACCTCTCT GGTCTTCAAC AACACTATCA TCAGGGAAAA CGTGGGGGAA
 201 GATGAACCAG CCGTGCAACT CGATGGAGCC GAGGGTGATG GACGATGACA
 251 TGCTCAAGCT GGCCGTCGGG GACCAGGGCC CCCAGGAGGA GGCCGGGCAG
 301 CTGGCCAAGC AGGAGGGCAT CCTCTTCAAG GATGTCCTGT CCCTGCAGCT
 351 GGACTTTCGG AACATCCTCC GCATAGACAA CCTCTGGCAG TTTGAGAACT
 401 TGAGGAAGCT GCAGCTGGAC AATAACATCA TTGAGAAGAT CGAGGGCCTG
 451 GAGAACCTCG CACACCTGGT CTGGCTGGAT CTGTCTTTCA ACAACATTGA
 501 GACCATCGAG GGGCTGGACA CACTGGTGAA CCTGGAGGAC CTGAGCTTGT
 551 TCAACAACCG GATCTCCAAG ATCGACTCCC TGGACGCCCT CGTCAAGCTG
 601 CAGGTGTTGT CGCTGGGCAA CAACCGGATT GACAACATGA TGAACATCAT
 651 CTACCTCCGG CGGTTCAAGT GCCTGCGGAC GCTCAGCCTC TCTAGGAACC
 701 CTATCTCTGA GGCAGAGGAT TACAAGATGT TCATCTGTGC CTACCTTCCT
 751 GACCTCATGT ACCTGGACTA CCGGCGCATT GATGACCACA CAGCAAGTGT
 801 CTCCCTCTCA GTCTCCCAGC CCTGTGAGAC AGATTCCTCA AGCCCCCAGG
 851 TTTCTTGGAA AAGGGGCATT GAAGAGTAGC TTCCCCTGCC CACAACTAGG
 901 AGAGAAAGGG CAGCTCCCTC TTCCTAATCC CTTTACCTGA CTCTGTCAGA
 951 GTGATTCCAG CAGCACCCTT GTAAGTACTG TTTTGTGTGC GTTCCCAGGG
1001 GCCAGGCCTC TTCCACACAC TGTCCCAGGG CCACCTCACA GCCATCCTGC
1051 ACTGTCTAGT TTTCCAGATG AAGAAGCTGA GGAGGGCTGG GAGCAGTGGC
1101 TCACGCCTGT AATCCCAGCA CTTTGAGAGG CTGAGGCGGG AGGATCGCTT
1151 GAGCCAAGGA GTTCAAGACC AGCCTGGGCA ACATAGGGAG ACCCCATCTC
1201 TACAGAAACT ACCAAAATTA GCCAGGTGTG GTGGCACACA CCAGTAATCC
1251 TGGCTACTCA CAAGGCCGAG GTAGAAGAAT CGCTTGAGAC TAGGAGTTTG
1301 AGGCTGCAGT GAACTAAGAA GATGCCATTG CACTCCAGCC TGGGCAACAG
1351 AGTGAAAAAA TTAAAAAATT AGAAAAGAAA AGAAGTTGAG GAGGCCCAAG
1401 GAGGGCAAGC AGCCAGGATC ACTGGCTCAA GGCCAAGCCA GGATTCACCC
1451 TAAGTTGGTG TCATCCCAGG AGCAATATTA ACAGCTGAGC TCCAGAGGGA
1501 ACCAGGCCAT CAGAGGCTCA GGCCTGGCTC TCAGGGGCAG AGTCAGGGCT
1551 GGAGGTAGAG ACCTGAGTGT CATCTGAGGA TTGCCAATTG GCAGTAGTTG
1601 AAGCCATGGT ACAGGTGGGA TCACCTGGGG CACATGGAGT GAGCTGGGGG
1651 ACGGGGACTA AGTTCTAGAG GTGCCAGCAT TCCTGGCCAG GTACAGGGGG
1701 ATGAGCCAGT GCGGTGGAGA GAGCCAAGGG CCAGACCCTC GTGACCAGCC
1751 CTATGGCCTC ACTCTACCTC TGTCCTGTTG TCCTCCTTCC CTAAAAGAGG
1801 GCCAGAAGGC CTGCTGAGGG CTGTTGGGAG TGAGAGAGCA AGTCCTCTGT
1851 GGAGAACACC CAGTCTGGGG CGAGGGGAGC GCTCCATTGC TGTGGCTCCT
1901 GCCCTGGAGA TGGCCCCGGG AACCCCAGCC TGCCACGCTG CCTTCCGCTC
1951 CTCCTGGTCT TTCCCTGATT TCCCTGCGCT CACAAAAACC TGGTGAGGGT
2001 CATCAGGAGA TGGGCATTCT CATCCACGAG ACCTCATGGC TTTCACAGCC
2051 TTCATGCAGG CCCCTGTGCA ACACCCCTGC CCATGCGCGG GAGGCTGCAG
2101 CATGGCAGAG GCGGCATGGC AGAGGCGGTG TGGCTCGGAG GAACCTCTGG
2151 TAACAATGCC ACTCCCGTTC CCTGGTCAGA AAAAGCTTGC GGAGGCTAAG
2201 CACCAGTACA GCATCGACGA GCTGAAGCAC CAGGAGAACC TGATGCAGGC
2251 CCAGCTGGAG GACGAGCAGG CGCAGCGGGA GGAGCTAGAG AAGCACAAGA
2301 CTGCGTTTGT GGAACACCTG AATGGCTCCT TCCTGTTTGA CAGCATGTAC
2351 GCTGAGGACT CAGAGGGCAA CAATCTGTCC TACCTGCCTG GTGTCGGTGA
2401 GCTCCTTGAG ACCTACAAGG ACAAGTTTGT CATCATCTGC GTGAATATTT
2451 TTGAGTATGG CCTGAAACAG CAGGAGAAGC GGAAAACAGA GCTTGACACC
2501 TTCAGTGAAT GTGTCCGTGA GGCCATCCAG GAAAACCAGG AGCAGGGCAA
```

```
2551 ACGCAAGATT GCCAAATTCG AGGAGAAGCA CTTGTCGAGT TTAAGTGCCA
2601 TTCGAGAGGA GTTGGAACTG CCCAACATTG AGAAGATGAT CCTAGAATGC
2651 AGTGCTGACA TCAGTGAGTT GTTCGATGCG CTCATGACGC TGGAGATGCA
2701 GCTGGTGGAG CAGCTGGAGG TAAGGCTGGG CCCTGGGCAC AAGTGCCAGA
2751 ATCTGGCGAT GCAGCTGCAC ATCCATAGGT GAACTGTAGC CTTCATGGGC
2801 ACGCCTCTGC TGGAAACGTC CAGCACGACT CAGCGTGGCA GGCTGTAGCT
2851 TTCTTGCTCA TCAGTCCTGT TTGCTTTTAT TACATTTTAA TCATTTACAT
2901 TGGAAGTGAT TCTTGTGGAA AATGAGAGGT GAGCTCATTC TTCTGAAATG
2951 GTCCCCCTAT CCTGGAAGTC AGTGGGGAGA GGTTTTTGAT TAGACCCCTG
3001 GAGCTATCCG GGTACTCTAA AGGCAAAGCG CACCCCCACT TGGGGACCAA
3051 ACAAAGACCC CTCCGCATTG CAGCCTGCAG TTGCCGCTTC TCAGGTGACG
3101 TGAGGAGGCT GCAACTCAGC ACTAAGTAGT GAAAATGAAA AGCGCCGCTG
3151 TCTGAAATTC ATTAGCAGCC AGAGTATGTG TTACAAGGCA GCGGAGGCTG
3201 GGAGTCTGAA GTGGTGTGAT GAATTGAACC TCATCGGATG CTGCTGTGGC
3251 TGGGCCAAGT GATAGCACCT AATCAATTCC TCACACGTCA AGTGACACCT
3301 CAGACATGGG ATAGATTTCC CCATCACATC ACAGGGCAGG TGCTCCCTCC
3351 CTGCTGGAGA GCACAGGCAC TGCAGAAGCA GCGCACAGTG CCAGGGGCGA
3401 GTGAGGCAGC AGCTCCCAGC CTTTTCAGGC ACGGAGATTG CCTTTCAACA
3451 TCCAAACATT TCCCAGAACC CATGTGCCAT CCTACTTGTA TTACTGGTGG
3501 CCAGAAAGCC ACAAGCGCAA TCATGCTTTT CAATGACCCT ATTTTTATTC
3551 ACGAGAACAG CACATACATG TGTTTGAAAA TTATGTGAGG TGCTCACTCT
3601 GCAGACAGTA CTCACATTCC TATAGATTCC ACCCCTGCCC ACCTTGCAGC
3651 CCCTGGAGTC TATAGCAGAT GGGAGTGGGG CACTCCGAGA GTGGCAGGCC
3701 TGGAGATCAC ATCTTCCATT GTTCCTTCAA TCAACACTAA CTCCCATTTG
3751 GGCCTTAGGT GCCTTGCTAA GCACCACAAA ACAGCAACTA ACTGAAAGAG
3801 ATCTGGAGTG CCAGCCCGCT CCTACTGAGG GCCTCCTCTC TGTCAGGCAC
3851 CTTGCAAAGC ATTTTGTGTG AAGTGACTCA TTTAACCTCA CCACAACGCC
3901 ACAACGCAGG GATTATGCAG GTAACCTATT TCCCAGATGA GGAAGATAAG
3951 GCCCAAGGAG GTGAAATGCC TTTCCCAGAG TTACACAGAG TGCTGGAGCT
4001 GGGAATACTG ACCCAGGCAG TCTAGCTCTT AACAGCTCAC TCCACTGTTT
4051 CCCTGGAGGT GATGCACAGA TGTCACTGGG AAACCCAAAG GAGAGGGGGT
4101 TGGCTGTGTG TGTGTGTGTT GGGCAGGCAG GTAAGGGGAG TAAGACCAGG
4151 ACAAGTGTTC CTGGCAAAGT TCCGGTGACA GCATTAAACA TTCAGATGGT
4201 GAGGGAGTTA ATATGGTTGG AGAACAACAA CTTTAGAGAG AGCAGAGGGG
4251 TCAGTTCACA ACCATCTGCT CAGGAGGGTC AAGATGGGTG GTCTTTATGC
4301 TGAAGGTCTG TGATTAGAGG AGCTGGTTGC TAAATTTTGA GGAGTACCTT
4351 TTGCTCTGTG CTGGACATCT AAATATGCAT GTTAACTGTG TTCTTTAACA
4401 TTTCCAGGAG ACTATAAACA TGTTTGAAAG GAACATTGTT GACATGGTAG
4451 GACTGTTTAT CGAAAATGTC CAAAGCCTAT ATCCTTTCTG TGATGACCTT
4501 CCCCATGGGG AGGTGCTACA GAGCCCCTGG GCTTGTCCCG GCCTCTGGAC
4551 AAAAGAATGT TCCACAGGGT CTGAGGAGGT TTCCCGACCC TCAGAACAAT
4601 GATGGCCTGG TTAGAGCTGT GGTTTGGATG CCCAGAGGGA CAACATCCAA
4651 ACTGTTTGCA GTAGGCTCCC AGCATGATTG TTCTCATATG AGTGATGTTC
4701 ACTAGGAAAT GACGCCCCCT GTGTTGCAGG CAAGCACACT CTGGGGTTGA
4751 GGCAACCCCC ACGTGGAAGA CACTATAAGG AGTACATCAG GTGAAATGTT
4801 AGGGTGAGGA GCCAACATCG GAGCATGGCC AACCCTTCTT CCACCCGAAC
4851 TCAGGGCACT CCACATGGGG CAAACTGCTG TGCTCCAGCT AGCAGCAGCC
4901 CTGTGGTCCT GCCCTCCTGG GGCTCACAGT CCCTCAGGGA GACAAGTTGT
4951 AGAGGCAACA AGTGGTGCCA AATGCACAGG GTGAGAAGCA GTTAACCCAG
5001 AGGCCAGGAG CCTCCATGCA GGAGGGAGAG AAGAGTGTGA TGGCAGGGGC
5051 CGAGGGTCCG TCCGAGGTGT GGGGCAGGGG CAGGGAGTCG AGGAAGGCCC
5101 AGGGTTCGGA GCTTGTGAGT GGACGGTGCT GCCAGCCAGA ATTTCCGAGC
5151 TCGCCTTGGG CCCTTAAAGT CTGTCTCCCG CCGTCTGAGA GCATCAGGGA
5201 CGCGCCGGGC CTGCTCCTCC CGGGCCTTTG CTTAACTCGG GGCTGCACGA
5251 TGGCTCAGTG CCGGGACCTG GAGAATCACC ACCACGAGAA GCTCCTGGAG
5301 ATCTCTATCA GCACCCTGGA GAAGATTGTC GAGGGCGACC TGGACGAGGA
5351 CCTGCCTAAC GACCTGCGCG CGCTTTTTGT CGATAAAGAT ACGATTGTTA
5401 ATGCTGTCGG GGCATCGCAC GACATCCACC TCCTGAAGAT TGACAATCGA
5451 GAAGATGAGC TGGTGACCAG AATCAACTCT TGGTGTACAC GTTTAATAGA
5501 CAGGATTCAC AAGGATGAGA TCATGAGGAA CCGCAAGCGC GTGAAGGAGA
5551 TCAATCAGTA CATCGACCAC ATGCAGAGCG AACTGGACAA CCTGGAATGT
5601 GGCGACATCC TAGACTAGAT GAATGTCAGC CACAGGAGCT TCTTCAAAAC
5651 ATAGCACCAG CCCCAGCCAG GAGAAGGAAG TGCACACGCC TCACCCGCAC
5701 CTCTAGAGAG TTGCTGGGCA TCTCTCAACC GCGATCCCCA ACACCATTCT
5751 TCCCCCACCC CTGGAAAAAC TTCCAAAAGT AGAGAAAATA AAGGACTCAT
5801 TTCACAAAAA AAAAAAAAAA AA
```

BLAST Results

Entry HS1292248 from database EMBL:
human STS SHGC-53917.
Score = 874, P = 3.3e-33, identities = 180/185

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 202 bp to 876 bp; peptide length: 225
Category: similarity to known protein

```
  1 MNQPCNSMEP RVMDDDMLKL AVGDQGPQEE AGQLAKQEGI LFKDVLSLQL
 51 DFRNILRIDN LWQFENLRKL QLDNNIIEKI EGLENLAHLV WLDLSFNNIE
101 TIEGLDTLVN LEDLSLFNNR ISKIDSLDAL VKLQVLSLGN NRIDNMMNII
151 YLRRFKCLRT LSLSRNPISE AEDYKMFICA YLPDLMYLDY RRIDDHTASV
201 SLSVSQPCET DSSSPQVSWK RGIEE
```

BLASTP hits

Entry S68209 from database PIR:
sds22 protein homolog - human >TREMBL:HSSDS22MR_1 gene: "sds22"; product: "yeast sds22 homolog"; H.sapiens sds22-like mRNA
Score = 234, P = 1.2e-19, identities = 61/143, positives = 93/143

Entry A38439 from database PIR:
suppressor protein sds22(+) - fission yeast (Schizosaccharomyces pombe) >TREMBL:SPSDS22_1 gene: "sds22+"; S.pombe sds22+ gene, complete cds.
Score = 208, P = 5.6e-17, identities = 52/127, positives = 71/127

Entry S43988 from database PIR:
protein suppressor sds22 - fission yeast (Schizosaccharomyces pombe) >SWISSPROT:SD22_SCHPO PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT SDS22. >TREMBL:SPAC4A8_12 gene: "sds22"; product: "phosphatases ppl regulatory subunit"; S.pombe chromosome I cosmid c4A8.
Score = 208, P = 8.5e-17, identities = 52/127, positives = 71/127

Entry CEK10D2_5 from database TREMBL:
gene: "K10D2.1"; Caenorhabditis elegans cosmid K10D2.
Score = 214, P = 3.6e-16, identities = 50/125, positives = 75/125

Alert BLASTP hits for DKFZphutel_20m11, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphutel_20m11, frame 1

Report for DKFZphutel_20m11.1

```
[LENGTH]        225
[MW]            25955.87
[pI]            4.63
[HOMOL]         PIR:S68209 sds22 protein homolog - human 1e-18
[FUNCAT]        03.22 cell cycle control and mitosis  [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]        30.10 nuclear organization    [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]        06.07 protein modification (glycolsylation, acylation, myristylation,
palmitylation, farnesylation and processing)       [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]        30.05 organization of centrosome       [S. cerevisiae, YOR373w] 2e-06
[FUNCAT]        01.03.10 metabolism of cyclic and unusual nucleotides       [S. cerevisiae,
YJL005w] 3e-05
[FUNCAT]        03.10 sporulation and germination     [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]        30.02 organization of plasma membrane         [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]        10.04.03 second messenger formation  [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]        04.07 rna transport   [S. cerevisiae, YPL169c] 9e-04
[FUNCAT]        04.05.01.04 transcriptional control  [S. cerevisiae, YCR065w] 9e-04
[EC]            4.6.1.1 Adenylate cyclase 2e-06
[PIRKW]         nucleus 5e-16
[PIRKW]         duplication 2e-06
[PIRKW]         tandem repeat 2e-06
[PIRKW]         cAMP biosynthesis 2e-06
[PIRKW]         glycoprotein 2e-06
[PIRKW]         phosphorus-oxygen lyase 2e-06
[SUPFAM]        leucine-rich alpha-2-glycoprotein repeat homology 5e-16
[SUPFAM]        fibromodulin 3e-07
[SUPFAM]        yeast adenylate cyclase catalytic domain homology 2e-06
[SUPFAM]        yeast adenylate cyclase 2e-06
[PROSITE]       CK2_PHOSPHO_SITE        2
[PROSITE]       PKC_PHOSPHO_SITE        1
```

[KW] All_Alpha

```
SEQ   MNQPCNSMEPRVMDDDMLKLAVGDQGPQEEAGQLAKQEGILFKDVLSLQLDFRNILRIDN
PRD   cccccccccccccchhhhhhcccccchhhhhhhhhhhchhhhhhhhhhcccccccccccc

SEQ   LWQFENLRKLQLDNNIIEKIEGLENLAHLVWLDLSFNNIETIEGLDTLVNLEDLSLFNNR
PRD   hhhhhhhhhhhhcccccccccccchhhhhhhhcccccccccccccchhhhhhhhhccccc

SEQ   ISKIDSLDALVKLQVLSLGNNRIDNMMNIIYLRRFKCLRTLSLSRNPISEAEDYKMFICA
PRD   cccchhhhhhhhhhhhccccccccccccccchhhhhhhhhccccccccccchhhhhhhhh

SEQ   YLPDLMYLDYRRIDDHTASVSLSVSQPCETDSSSPQVSWKRGIEE
PRD   hhccccccccccccchhhhhhhhcccccccccccccccccccc
```

Prosite for DKFZphutel_20m11.1

PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	169->173	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphutel_20m11.1)

DKFZphutel_20m24
----

group: metabolism

DKFZphutel_20m24 encodes a novel 611 amino acid protein with similarity to a hypothetical C.elegans protein and to yeast Alg9 protein.

This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2.

The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

strong similarity to S.cerevisiae Alg9p

complete cDNA, complete cds, potential start at Bp 23, few EST hits
Alg9 is involved in the assembly of the core oligosaccharide
Glc3Man9GlcNAc2
HSAC381 corresponding genomic DNA (2 exons)
HSB8954 corresponding genomic DNA (1 exon)

Sequenced by AGOWA

Locus: /map="11"

Insert length: 1986 bp
Poly A stretch at pos. 1966, polyadenylation signal at pos. 1949

```
   1 TTCTTTTTTC CCCAGGCTTG CCATGGCTAG TCGAGGGGCT CGGCAGCGCC
  51 TGAAGGGCAG CGGGGCCAGC AGTGGGGATA CGGCCCCGGC TGCGGACAAG
 101 CTGCGGGAGC TGCTGGGCAG CCGAGAGGCG GGCGGCGCGG AGCACCGGAC
 151 CGAGTTATCT GGGAACAAAG CAGGACAAGT CTGGGCACCT GAAGGATCTA
 201 CTGCTTTCAA GTGTCTGCTT TCAGCAAGGT TATGTGCTGC TCTCCTGAGC
 251 AACATCTCTG ACTGTGATGA AACATTCAAC TACTGGGAGC CAACACACTA
 301 CCTCATCTAT GGGGAAGGGT TTCAGACTTG GGAATATTCC CCAGCATATG
 351 CCATTCGCTC CTATGCTTAC CTGTTGCTTC ATGCCTGGCC AGCTGCATTT
 401 CATGCAAGAA TTCTACAAAC TAATAAGATT CTTGTGTTTT ACTTTTTGCG
 451 ATGTCTTCTG GCTTTTGTGA GCTGTATTTG TGAACTTTAC TTTTACAAGG
 501 CTGTGTGCAA GAAGTTTGGG TTGCACGTGA GTCGAATGAT GCTAGCCTTC
 551 TTGGTTCTCA GCACTGGCAT GTTTTGCTCA TCATCAGCAT TCCTTCCTAG
 601 TAGCTTCTGT ATGTACACTA CGTTGATAGC CATGACTGGA TGGTATATGG
 651 ACAAGACTTC CATTGCTGTG CTGGGAGTAG CAGCTGGGGC TATCTTAGGC
 701 TGGCCATTCA GTGCAGCTCT TGGTTTACCC ATTGCCTTTG ATTTGCTGGT
 751 CATGAAACAC AGGTGGAAGA GTTTCTTTCA TTGGTCGCTG ATGGCCCTCA
 801 TACTATTTCT GGTGCCTGTG GTGGTCATTG ACAGCTACTA TTATGGGAAG
 851 TTGGTGATTG CACCACTCAA CATTGTTTTG TATAATGTCT TTACTCCTCA
 901 TGGACCTGAT CTTTATGGTA CAGAACCCTG GTATTTCTAT TTAATTAATG
 951 GATTTCTGAA TTTCAATGTA GCCTTTGCTT TGGCTCTCCT AGTCCTACCA
1001 CTGACTTCTC TTATGGAATA CCTGCTGCAG AGATTTCATG TTCAGAATTT
1051 AGGCCACCCG TATTGGCTTA CCTTGGCTCC AATGTATATT TGGTTTATAA
1101 TTTTCTTCAT CCAGCCTCAC AAAGAGGAGA GATTTCTTTT CCCTGTGTAT
1151 CCACTTATAT GTCTCTGTGG CGCTGTGGCT CTCTCTGCAC TTCAGAAATG
1201 TTACCACTTT GTGTTTCAAC GATATCGCCT GGAGCACTAT ACTGTGACAT
1251 CGAATTGGCT GGCATTAGGA ACTGTCTTCC TGTTTGGGCT CTTGTCATTT
1301 TCTCGCTCTG TGGCACTGTT CAGAGGATAT CACGGGCCCC TTGATTTGTA
1351 TCCAGAATTT TACCGAATTG CTACAGACCC AACCATCCAC ACTGTCCCAG
1401 AAGGCAGACC TGTGAATGTC TGTGTGGGAA AAGAGTGGTA TCGATTTCCC
1451 AGCAGCTTCC TTCTTCCTGA CAATTGGCAG CTTCAGTTCA TTCCATCAGA
1501 GTTCAGAGGT CAGTTACCAA AACCTTTTGC AGAAGGACCT CTGGCCACCC
1551 GGATTGTTCC TACTGACATG AATGACCAGA ATCTAGAAGA GCCATCCAGA
1601 TATATTGATA TCAGTAAATG CCATTATTTA GTGGATTTGG ACACCATGAG
1651 AGAAACACCC CGGGAGCCAA AATATTCATC CAATAAAGAA GAATGGATCA
1701 GCTTGGCCTA TAGACCATTC CTTGATGCTT CTAGATCTTC AAAGCTGCTG
1751 CGGGCATTCT ATGTCCCCTT CCTGTCAGAT CAGTATACAG TGTACGTAAA
1801 CTACACCATC CTCAAACCCC GGAAAGCAAA GCAAATCAGG AAGAAAAGTG
1851 GAGGTTAGCA ACACACCTGT GGCCCCAAAG GACAACCATC TTGTTAACTA
1901 TTGATTCCAG TGACCTGACT CCCTGCAAGT CATCGCCTGT AACATTTGTA
1951 ATAAAGGTCT TCTGACATGA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HSAC381 from database EMBL:
Homo sapiens chromosome 11 pac pDJ159o1, complete sequence.
Length = 42,771

Entry HSB8954 from database EMBL:

cSRL-50A3-u cSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone cSRL-50A3.
Length = 601

Medline entries

96293493:
Stepwise assembly of the lipid-linked oligosaccharide in the endoplasmic reticulum of Saccharomyces cerevisiae: identification of the ALG9 gene encoding a putative mannosyl transferase.

Peptide information for frame 2

ORF from 23 bp to 1855 bp; peptide length: 611
Category: strong similarity to known protein

```
  1 MASRGARQRL KGSGASSGDT APAADKLREL LGSREAGGAE HRTELSGNKA
 51 GQVWAPEGST AFKCLLSARL CAALLSNISD CDETFNYWEP THYLIYGEGF
101 QTWEYSPAYA IRSYAYLLLH AWPAAFHARI LQTNKILVFY FLRCLLAFVS
151 CICELYFYKA VCKKFGLHVS RMMLAFLVLS TGMFCSSSAF LPSSFCMYTT
201 LIAMTGWYMD KTSIAVLGVA AGAILGWPFS AALGLPIAFD LLVMKHRWKS
251 FFHWSLMALI LFLVPVVVID SYYYGKLVIA PLNIVLYNVF TPHGPDLYGT
301 EPWYFYLING FLNFNVAFAL ALLVLPLTSL MEYLLQRFHV QNLGHPYWLT
351 LAPMYIWFII FFIQPHKEER FLFPVYPLIC LCGAVALSAL QKCYHFVFQR
401 YRLEHYTVTS NWLALGTVFL FGLLSFSRSV ALFRGYHGPL DLYPEFYRIA
451 TDPTIHTVPE GRPVNVCVGK EWYRFPSSFL LPDNWQLQFI PSEFRGQLPK
501 PFAEGPLATR IVPTDMNDQN LEEPSRYIDI SKCHYLVDLD TMRETPREPK
551 YSSNKEEWIS LAYRPFLDAS RSSKLLRAFY VPFLSDQYTV YVNYTILKPR
601 KAKQIRKKSG G
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_20m24, frame 2

SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (Saccharomyces cerevisiae), N = 1, Score = 533, P = 2.3e-51

SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (Saccharomyces cerevisiae), N = 1, Score = 533, P = 2.3e-51

>SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II.
Length = 653

HSPs:

Score = 957 (143.6 bits), Expect = 2.7e-96, P = 2.7e-96
Identities = 206/514 (40%), Positives = 296/514 (57%)

```
Query:   48 NKAGQVWAPEGSTAFKCLLSARLCAALLSNISDCDETFNYWEPTHYLIYGEGFQTWEYSP 107
            N   W      + FK LLS R+  A+   I+DCDE +NYWEP H  +YGEGFQTWEYSP
Sbjct:   43 NNPDNDWPFSFGSVFKMLLSIRISGAIWGIINDCDEVYNYWEPLHLFLYGEGFQTWEYSP 102

Query:  108 AYAIRSYAYLLLHAWPAAFHARILQTNKILVFYFLRCLLAFVSCICELYFYKAVCKKFGL 167
             YAIRSY Y+ LH  PA+  A +     KI+VF  +R  +     + E Y + A+CKK  +
Sbjct:  103 VYAIRSYFYIYLHYIPASLFANLFGDTKIVVFTLIRLTIGLFCLLGEYYAFDAICKKINI 162

Query:  168 HVSRMMLAFLVLSTGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGW 227
              R  + F + S+GMF +S+AF+PSSFCM  T  +  +  +  +  +  VA   ++GW
Sbjct:  163 ATGRFFILFSIFSSGMFLASTAFVPSSFCMAITFYILGAYLNENWTAGIFCVAFSTMVGW 222

Query:  228 PFSAALGLPIAFDLLVMKHRWKSFFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLY 287
```

```
                PFSA LGLPI  D+L++K     F   SL+  +        V+ DS+Y+GK V+APLNI LY
Sbjct:     223 PFSAVLGLPIVADMLLLKGLRIRFILTSLVIGLCIGGVQVITDSHYFGKTVLAPLNIFLY 282

Query:     288 NVFTPHGPDLYGTEPWYFYLINGFLNFNVAFALALLVLPLTSLMEYLLQRFHVQNLGHPY 347
               NV +   GP LYG EP   FY+ N F N+N+     A     PL+   + Y  + +  Q+
Sbjct:     283 NVVSGPGPSLYGEEPLSFYIKNLFNNWNIVIFAAPFGFPLS--LAYFTKVWMSQDRNVAL 340

Query:     348 WLTLAPMYI-------WFIIFFIQPHKEERFLFPVYPLICLCGAVALSALQKCYHFVFQR 400
               +   AP+ +         W +IF  Q HKEERFLFP+YP I    A+AL A  +       ++
Sbjct:     341 YQRFAPIILLAVTTAAWLLIFGSQAHKEERFLFPIYPFIAFFAALALDATNR---LCLKK 397

Query:     401 YRLEHYTVTSNWLALGTVFLFGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPE 460
                ++       N L++  +  F +LS SR+ ++    Y    +++Y          T+ T +
Sbjct:     398 LGMD------NILSILFILCFAILSASRTYSIHNNYGSHVEIYRSLNAELTNRT-NFKNF 450

Query:     461 GRPVNVCVGKEWYRFPSSFLLPDNW------QLQFIPSEFRGQLPKPFAEGPL---ATRI 511
                 P+ VCVGKEW+RFPSSF +P              +++FI SEFRG LPKPF +          TR
Sbjct:     451 HDPIRVCVGKEWHRFPSSFFIPQTVSDGKKVEMRFIQSEFRGLLPKPFLKSDKLVEVTRH 510

Query:     512 VPTDMNDQNLEEPSRYIDISKCHYLVDLDTMRETPREPKYSSNKEEW 558
               +PT+MN+ N EE SRY+D+  C Y+VD+D M ++ REP +   ++ +
Sbjct:     511 IPTEMNNLNQEEISRYVDLDSCDYVVDVD-MPQSDREPDFRKMRQNY 556
```

Pedant information for DKFZphute1_20m24, frame 2

Report for DKFZphute1_20m24.2

```
[LENGTH]     611
[MW]         69863.78
[pI]         8.91
[HOMOL]      SWISSPROT:YTH3_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II. 2e-
93
[FUNCAT]     09.01 biogenesis of cell wall        [S. cerevisiae, YNL219c] 4e-69
[FUNCAT]     01.06.01 lipid, fatty-acid and sterol biosynthesis   [S. cerevisiae, YNL219c]
4e-69
[FUNCAT]     01.05.01 carbohydrate utilization      [S. cerevisiae, YNL219c] 4e-69
[PIRKW]      glycosyltransferase 9e-68
[PIRKW]      transmembrane protein 9e-68
[PIRKW]      hexosyltransferase 9e-68
[PROSITE]    MYRISTYL         9
[PROSITE]    CAMP_PHOSPHO_SITE      1
[PROSITE]    CK2_PHOSPHO_SITE       7
[PROSITE]    PKC_PHOSPHO_SITE       6
[PROSITE]    ASN_GLYCOSYLATION      2
[KW]         TRANSMEMBRANE   7
[KW]         LOW_COMPLEXITY      6.71 %
```

```
SEQ    MASRGARQRLKGSGASSGDTAPAADKLRELLGSREAGGAEHRTELSGNKAGQVWAPEGST
SEG    ............................................................
PRD    ccchhhhhhcccccccccccchhhhhhhhccccccccccccceeeccccccccccccccch
MEM    ......................................................MMMMMM

SEQ    AFKCLLSARLCAALLSNISDCDETFNYWEPTHYLIYGEGFQTWEYSPAYAIRSYAYLLLH
SEG    ...xxxxxxxxxxxxx............................................
PRD    hhhhhhhhhhhhhhhhhhccccceeeccccceeeeeccccceeeccchhhhhhhhhhhhc
MEM    MMMMMMMMMMMMMMMMM..........................................M

SEQ    AWPAAFHARILQTNKILVFYFLRCLLAFVSCICELYFYKAVCKKFGLHVSRMMLAFLVLS
SEG    ............................................................
PRD    cchhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc
MEM    MMMMMMMMMMMMMMMMMMMMMMMMMMM.................................

SEQ    TGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGWPFSAALGLPIAFD
SEG    .................................xxxxxxxxxxxxx..............
PRD    cceeeecccccchhhhhhhhhhhhhcccccccceeeeehhhhhccceeeeeeecchhhhh
MEM    ..............................................MMMMMMMMMMMMMM

SEQ    LLVMKHRWKSFFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLYNVFTPHGPDLYGT
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhhheeeeeeeecccccccccccceeeeeeeeeccccccccc
MEM    MMMMMMM.MMMMMMMMMMMMMMMMMMMM................................

SEQ    EPWYFYLINGFLNFNVAFALALLVLPLTSLMEYLLQRFHVQNLGHPYWLTLAPMYIWFII
SEG    ..........xxxxxxxxxxxxxxx...................................
PRD    cceeeeeecccccchhhhhhhhhhhhhchhhhhhhhhhhhhcccccceeeehhhhhhhhh
MEM    .......MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.......................
```

```
SEQ    FFIQPHKEERFLFPVYPLICLCGAVALSALQKCYHFVFQRYRLEHYTVTSNWLALGTVFL
SEG    ............................................................
PRD    hhcccchhhhhcccceeehhhhhhhhhhhhhhhhhhhhhhhhhheeeeccchhhhhhhee
MEM    ............MMMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMM.

SEQ    FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCVGKEWYRFPSSFL
SEG    ............................................................
PRD    eehhhhhhheeeccccccccccccceeeecccccceeecccceeeeeecccccccccccc
MEM    ............................................................

SEQ    LPDNWQLQFIPSEFRGQLPKPFAEGPLATRIVPTDMNDQNLEEPSRYIDISKCHYLVDLD
SEG    ............................................................
PRD    ccccceeeeccccccccccccccccceeeecccccccccccccccceeeeeeceeeeecc
MEM    ............................................................

SEQ    TMRETPREPKYSSNKEEWISLAYRPFLDASRSSKLLRAFYVPFLSDQYTVYVNYTILKPR
SEG    ............................................................
PRD    cccccccccccchhhhhhhhhhhhhhhhhhheeeeeeeeecceeeeeeeeeeeeeecccc
MEM    ............................................................

SEQ    KAKQIRKKSGG
SEG    ...........
PRD    hhhhhhccccc
MEM    ...........
```

Prosite for DKFZphutel_20m24.2

PS00001	77->81	ASN_GLYCOSYLATION	PDOC00001
PS00001	593->597	ASN_GLYCOSYLATION	PDOC00001
PS00004	606->610	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	67->70	PKC_PHOSPHO_SITE	PDOC00005
PS00005	133->136	PKC_PHOSPHO_SITE	PDOC00005
PS00005	541->544	PKC_PHOSPHO_SITE	PDOC00005
PS00005	545->548	PKC_PHOSPHO_SITE	PDOC00005
PS00005	553->556	PKC_PHOSPHO_SITE	PDOC00005
PS00005	572->575	PKC_PHOSPHO_SITE	PDOC00005
PS00006	16->20	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	329->333	CK2_PHOSPHO_SITE	PDOC00006
PS00006	457->461	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00006	545->549	CK2_PHOSPHO_SITE	PDOC00006
PS00006	553->557	CK2_PHOSPHO_SITE	PDOC00006
PS00008	12->18	MYRISTYL	PDOC00008
PS00008	14->20	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	47->53	MYRISTYL	PDOC00008
PS00008	166->172	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	218->224	MYRISTYL	PDOC00008
PS00008	222->228	MYRISTYL	PDOC00008
PS00008	234->240	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphutel_20m24.2)

DKFZphutel_21d15
----

group: uterus derived

DKFZphutel_21d15 encodes a novel 191 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /chromosome="3"

Insert length: 5292 bp
Poly A stretch at pos. 5273, polyadenylation signal at pos. 5252

```
   1 CTCCCACTAG TGTATGCCTT AATGGTGCCG CTCTTGTCCG CGTCTACGCT
  51 TGGGACCTTG GCTTCTGACT TGGAGAGTGT ACAGCTCTGC CCGACGGCAA
 101 CCCAGCTTGG GAAGAGAAGC CCCAGCGTGG GCTGGGGCTC AAGGCGCAGG
 151 AAGGCCGAGC CCGGCGCGGA CGCAGGCGGC TCCGGGCGGG CTCAGCACCC
 201 CCAGGCACCG TCTCCTAGTG ACCGCGGCGC TCGCGGGCCT GGCGGCCGTT
 251 GTCCGGGCGA CTGCGCAGCG CGGGCACCCC CGCGGCCCCT CCCCTGGGCG
 301 CGCGCGCGAC CTGGGTGCCA TGGCGGCAGC GGCGGTGACA GGCCAGCGGC
 351 CTGAGACCGC GGCGGCCGAG GAGGCCTCGA GGCCGCAGTG GGCGCCGCCA
 401 GACCACTGCC AGGCTCAGGC GGCGGCCGGG CTGGGCGACG GCGAGGACGC
 451 ACCGGTGCGT CCGCTGTGCA AGCCCCGCGG CATCTGCTCG CGCGCCTACT
 501 TCCTGGTGCT GATGGTGTTC GTGCACCTGT ACCTGGGTAA CGTGCTGGCG
 551 CTGCTGCTCT TCGTGCACTA CAGCAACGGC GACGAAAGCA GCGATCCCGG
 601 GCCCCAACAC CGTGCCCAGG GCCCCGGGCC CGAGCCCACC TTAGGTCCCC
 651 TCACCCGGCT GGAGGGCATC AAGGTGAGGA CCTCCCTGCC CCGCCGCGCT
 701 CCAGGCCCTG CACGGCTGAG CCCGAGAGGA CCGGCGCTCA GCCCGGGTCC
 751 CCACGCTGCC CCCGGCGCTG CTCTGCGTCG GTCCCGCGCG CTCCCACTCA
 801 CTCGCCTGCT GTCGCTCTCC GGGCCGGGGC GACTTGGCCC TTTTTGGGCA
 851 GCGCGGTCTG GCGCCCCAGC TGCCCGCTGT GCGCCTTTTC CTTAGGTGGG
 901 GCACGAGCGT AAGGTCCAGC TGGTCACCGA CAGGGATCAC TTCATCCGAA
 951 CCCTCAGCCT CAAGCCGCTG CTCTTCGAAA TCCCCGGCTT CCTGACTGAT
1001 GAAGAGTGTC GGCTCATCAT CCATCTGGCG CAGATGAAGG GGTTACAGCG
1051 CAGCCAGATC CTGCCTACTG AAGAGTATGA AGAGGCAATG AGCACTATGC
1101 AGGTCAGCCA GCTGGACCTC TTCCGGCTGC TGGACCAGAA CCGTGATGGG
1151 CACCTTCAGC TCCGTGAGGT TCTGGCCCAG ACTCGCCTGG GAAATGGATG
1201 GTGGATGACT CCAGAGAGCA TTCAGGAGAT GTACGCCGCG ATCAAGGCTG
1251 ACCCTGATGG TGACGGTGAG CTCACACCTC TGCACAGTCC TATCCCCGTG
1301 AGCCTCCTGC CCACTCCCAG GTGCACAATT TTGAAAACTT GGGCCCTTCC
1351 CCCACAGCCA GGCAGCCTCT CTGCACCCCT TTATAGTGGC CAGAGATGGG
1401 GAGGTGAAGA TCCAGCCTTG CTTTTTACCC CTGGGAAGTA GGCAGGCAGC
1451 CAGGCCCCCC GTTCCCCTTG GTGATGGTCT CGAGGGCAGT TCTTGGAGAC
1501 CCTTTTGATA ACATCAGGCA GAGTTGAGAG CCTGGGGACA GGAAGTAGGG
1551 CTGCTAGTTG GCAGAGAACA GAGTGGGTGG AGCAGGAGCA AGGCGACAGT
1601 GAGGCCAGCT AGAGCTTGGC TGTTTACCCT GCTCCATCCA TCTCTCCAGC
1651 CAGACACGAG GTCCACCCCA GCAGACAGCT TCCCTGGTCT AAGTGAGGTC
1701 TCCCTTGCCT TCCTCTTGTC CACCTGGAGT CATGCCGAAG CGCCTAAAAT
1751 GGTAGTGCTG CTACCTGTGC TAACTGCTGG GGAGGGGTGG GCAGGGAAGC
1801 TGTCATGCAA GTGGTGCCCC CTCTGGTAAT AACTCTCAGG AGGTTTCTGA
1851 GGTGTGGTCA TCACCCTCAT GCCCAAATTC TGGACCAAGA GAGGAAGATA
1901 CAGCAGTTAG AAAGGACTTG GAACAGTGGC TTTGCGGCTG GTGAACCAGA
1951 GTGAAGAATC TGGCCGTGAC CTGGCTGCCA CACTGCTATA GGCCCCAGAA
2001 CAGAGGTGGT GACAGTCTCA CAGCCCTTGA ATGTCCCCCA CCCTCAGAGG
2051 AATCTGGGCC AAAGAGTGGA AGGTGATGTC CTTGGGTCAG CCAGAATAAC
2101 ATGGAGCAAA GATACCAACT ACTCTTCCAG AACCCCAAGA GGGTAGAACC
2151 CCTGCTTAAT GGTTTGAGCA GGGACAGTGG AGAATGTTCT CATGAGAGGG
2201 GGTGGCCTGA CTTTCGTTGC TAAGTGGGCT GGTAACGCAG TAGGCAGGGC
2251 TGGCGAAGTA GGTTCCACCC AGGATGAAAC CTGGGGTCAT GAGGAACTCC
2301 CCGGGGGCTG GCCCTGCTTG CACCCTGGCG TATGTATGTA AGGCCCTGGA
2351 TGAGGCCCAG CACTGCCTGC TCTCTCCTCA CCCTCCACAG GCCGGAGAGT
2401 GGCCACCACT CTATATAGCC AGGCTGGAAG GCCAGGGTCC TGGCCATATG
2451 GCTCAAGCTT CCTTTGGAGA ACCTTCTCTG GCCACTCTAA TAGGGGGTGG
2501 GCCTCTTTCT TCTTAGGGCC AAATTAGGGC TTAAACTGAG AAAAGGAACT
2551 GCTCTGGGTC TTCCTGTAAG GCCTGATGTG ACAGAAACCA GGTTCATCTG
2601 ACCCAAAAGT CCAGGTGGGG GACAAGTGTA CAAGGCCCCT CAGTGCCTGA
2651 GGTCAGGGGC TGCTGCTGCC TTTGGGGTAG GTAGGGAAGT GCAGCCTGCC
2701 ACTGTTGCCT CCCAATATGG GCTTGGTGGG CATTGATGGT GGGTGCCCTG
2751 TGCAGGAGTG CTGAGTCTGC AGGAGTTCTC CAACATGGAC CTTCGGGACT
2801 TCCACAAGTA CATGAGGAGC CACAAGGCAG AGTCCAGTGA GCTGGTGCGG
```

```
2851 AACAGCCACC ATACCTGGCT CTACCAGGGT GAGGGTGCCC ACCACATCAT
2901 GCGTGCCATC CGCCAGAGGT GAGCACCTGA AGCTGTTCTC ACTGGAGCAG
2951 GGGGAGAAGA CTGGGCAGGG CCTCCACAGA AGTCCTTGTC TGGGGCCAAG
3001 AGGACAGAAT GGATTAACCC ATTTGGGATT AAGTTCCATT TGTTAGACCA
3051 GGATTGGGAC CCACTGAAAG ACAGGCAATT AACAAAGGCA AATTAGCCCT
3101 CCTTGCAGGC ACACAATGGG CAACTGGGGT TAGATAGAGA TTGAGCACTT
3151 CTTTCTGATT AGATAAATGA CCTCTTATCT TTGACCCCTT ATCTGACCCC
3201 GTCACAGCAG GAAAAGGGTT TTTAAATAAA CAACTTTCTT CCAGGGAGGA
3251 GGACCTCAGG ACTCCCCGCC CCCTTTATTT AGTGGAAATG TCAACATTTC
3301 CACATAGCAG GTGTCTCTGT CTTTGGCATC TGAGGGAGAA GGATCATCAT
3351 GAGTAACCCC CTCCTGCTCT TACAGGGCCA GTCTGAGATG GCTTAAGGGA
3401 CTTCCAGGGG AGGTGGGTAG GGGCAAAGCT TGTGGCAGGC CTAGGGTCCA
3451 CCTTGGCCAG CTCCTTCAGA TCACCACCTT GCCTGGGGCT GCCCAGCCAA
3501 ATGCCTGCTG CCCACCAGGG TGCTGCGCCT CACTCGCCTG TCGCCTGAGA
3551 TCGTGGAGCT CAGCGAGCCG CTGCAGGTTG TTCGATATGG TGAGGGGGGC
3601 CACTACCATG CCCACGTGGA CAGTGGGCCT GTGTACCCAG AGACCATCTG
3651 CTCCCATACC AAGCTGGTAG CCAACGAGTC TGTACCCTTC GAGACCTCCT
3701 GCCGGCAAGT ATCTCCCAAC TGGGGGCTGC CTTCAATCCT CAGACCAGGA
3751 ACACCCATGA CACAGGCACA GCCCTGCACT GTGGGCGTGC CCCTTGGCAT
3801 GGGGCCAGGA GATCACTGGG TTATCCCGGT TAGTGATGCC CTCACCTCTC
3851 CCCACAAGTT GTTTACCCAA TGGCTGGAAA GGGGTGGCTA CTGGTCATCG
3901 TGACCACTGG AGTCAACACA GACTGATGTA CCCACAGACA CCAAAACTTG
3951 CCCCCTGAGT TCTGAAGCAA GGGGCAAGGC TGGGCCCCTA GCTTGTCCTG
4001 CCCATTCCTC CAGGTGTTGA TCTTGATTCC ACTTAGAGAA GCTGAAGCTG
4051 TGCCTCCCTC CCCTGTCAAG CCAGTTCTTT CCTCTTCAGG TGGCTGTTCT
4101 GGCCCAGCCC CTTCCCATCC CCAAGGAGCC CTTCAGCGCG CCCTGTTGCT
4151 TCTGCTAGCC TACCTTTCCC TGCCAGGCCC TTGCTCAGGG CCATGGCATT
4201 TAACTAAGTG CACCTGTGAT CTTGGCCAAA AAACCATTGC AACTCACAGT
4251 AAGAGACTGG GTTTCGGGGA AGGAGGGGCT AGGGACATTT TGGCACTGGC
4301 CTGCCCTATT GTCTCCCATC CTAGTCTGTC CTGGTCCCTG GCAACAGGAA
4351 CCTGGGCAGC TTATCCTGCC CACAGGTAAG CCCCTGGGAG CATCCACAAC
4401 TGGGGACCTG CTCAGTGCCC CCCCTGCCTT ACAGCTACAT GACAGTGCTG
4451 TTTTATTTGA ACAACGTCAC TGGTGGGGGC GAGACTGTTT TCCCTGTAGC
4501 AGATAACAGA ACCTACGATG AAATGGTAAG GGTCAACTGG GCTATTACTC
4551 TTGTGGGCTG GCAGGGGCTT AGACAAGTGA AGTACACACC TCTCCAGGTC
4601 TAAGGATGTG GGCCCAAATT ATTCCTTGGG CATATCTGGT TGGTTTCCCT
4651 TTGGTCACCC TTGGCTGGCC TGGCCATAGA GTGGGGACAG GTTGAACACC
4701 CCACCACCCT GCTGCCCACA GAGTCTGATT CAGGATGACG TGGACCTCCG
4751 TGACACACGG AGGCACTGTG ACAAGGGAAA CCTGCGTGTC AAGCCCCAAC
4801 AGGGCACAGC AGTCTTCTGG TACAACTACC TGCCTGATGG GCAAGGTTGG
4851 GTGGGTGACG TAGACGACTA CTCGCTGCAC GGGGGCTGCC TGGTCACGCG
4901 CGGCACCAAG TGGATTGCCA ACAACTGGAT TAATGTGGAC CCCAGCCGAG
4951 CGCGGCAAGC GCTGTTCCAA CAGGAGATGG CCCGCCTTGC CCGAGAAGGG
5001 GGCACCGACT CACAGCCCGA GTGGGCTCTG GACCGGGCCT ACCGCGATGC
5051 GCGCGTGGAA CTCTGAGGGA AGAGTTAGCC CCGGTTCCCA GCCGCGGGTC
5101 GCCAGTTGCC CAAGATCAGG GGTCCGGCTG TCCTTCTGTC CTGCTGCAGA
5151 CTAAAGGTCT GGCCAATGTC TTGCCCCACC CCGCCAGCCG CGATACGGCG
5201 CAGTTCCTAT ATTCATGTTA TTTATTGTGT ACTGACTCCA TCTGCCCCGT
5251 CAAATAAAAA ACCACAAGGT TCGAAAAAAA AAAAAAAAAA GG
```

BLAST Results

Entry HSU64252 from database EMBL:
Human STS sequence NOTI-225.
Score = 959, P = 1.2e-36, identities = 195/199

Medline entries

No Medline entry

Peptide information for frame 1

ORF from the beginning to 351 bp; peptide length: 118
Category: questionable ORF
Classification: no clue

```
  1 LPLVYALMVP LLSASTLGTL ASDLESVQLC PTATQLGKRS PSVGWGSRRR
 51 KAEPGADAGG SGRAQHPQAP SPSDRGARGP GGRCPGDCAA RAPPRPLPWA
101 RARPGCHGGS GGDRPAA
```

BLASTP hits

```
No BLASTP hits available

            Alert BLASTP hits for DKFZphutel_21d15, frame 1

No Alert BLASTP hits found

                    Peptide information for frame 2
                    -------------------------------

ORF from 320 bp to 892 bp; peptide length: 191
Category: putative protein
Classification: no clue

     1 MAAAAVTGQR PETAAAEEAS RPQWAPPDHC QAQAAAGLGD GEDAPVRPLC
    51 KPRGICSRAY FLVLMVFVHL YLGNVLALLL FVHYSNGDES SDPGPQHRAQ
   101 GPGPEPTLGP LTRLEGIKVR TSLPRRAPGP ARLSPRGPAL SPGPHAAPGA
   151 ALRRSRALPL TRLLSLSGPG RLGPFWAARS GAPAARCAPF P


                              BLASTP hits

No BLASTP hits available

            Alert BLASTP hits for DKFZphutel_21d15, frame 2

PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1, N = 2,
Score = 106, P = 0.0067


>PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1
            Length = 1,298

  HSPs:

 Score = 106 (15.9 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03
 Identities = 36/103 (34%), Positives = 44/103 (42%)

Query:    87 GDESSDPGPQHRAQGPGPEPTLGPLTRLEGIKVRTSLPRRA-PGPARLS-PRGPALSPGP 144
             G +    PGP      G GP P    P T+  G      S  R   P PA  S P GP  +P
Sbjct:   726 GRKRKSPGPARPPGGGGPRP---PKTKKSGADAPGSDARAPLPAPAPPSTPPGPEPAPAQ 782

Query:   145 HAAPGAALRRSRALPLT-RLLSLSGPGRLGPFWAARSGAPAARCAP 189
              AAP AA  ++R  P+          GP  LG  W  +   P+   AP
Sbjct:   783 PAAPRAAAAQARPRPVAVSRRPAEGPDPLGG-WRRQPPGPSHTAAP 827

 Score = 40 (6.0 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03
 Identities = 8/21 (38%), Positives = 9/21 (42%)

Query:    28 DHCQAQAAAGLGDGEDAPVRP 48
             DH +   A G G     AP  P
Sbjct:   212 DHAREARAVGRGPSSAAPAAP 232

            Pedant information for DKFZphutel_21d15, frame 1
            ------------------------------------------------

                    Report for DKFZphutel_21d15.1


[LENGTH]        117
[MW]            11797.32
[pI]            10.68
[KW]            Irregular
[KW]            SIGNAL_PEPTIDE 22
[KW]            LOW_COMPLEXITY     38.46 %

SEQ    LPLVYALMVPLLSASTLGTLASDLESVQLCPTATQLGKRSPSVGWGSRRRKAEPGADAGG
SEG    .........xxxxxxxxxxxxx......................................
PRD    cccccccccccccccccccchhhhhhhhcccccccccccccccccccccccccccccccc

SEQ    SGRAQHPQAPSPSDRGARGPGGRCPGDCAARAPPRPLPWARARPGCHGGSGGDRPAA
SEG    ............xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.............
PRD    ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

(No Prosite data available for DKFZphutel_21d15.1)
```

(No Pfam data available for DKFZphutel_21d15.1)

```
Pedant information for DKFZphutel_21d15, frame 2
------------------------------------------------

Report for DKFZphutel_21d15.2

[LENGTH]     191
[MW]         19916.88
[pI]         10.43
[KW]         TRANSMEMBRANE  1
[KW]         LOW_COMPLEXITY    29.84 %

SEQ   MAAAAVTGQRPETAAAEEASRPQWAPPDHCQAQAAAGLGDGEDAPVRPLCKPRGICSRAY
SEG   ............................................................
PRD   ccceeeecccchhhhhhhhcccccchhhhhhhhcccccccccccccccccccccchhhh
MEM   ............................................................

SEQ   FLVLMVFVHLYLGNVLALLLFVHYSNGDESSDPGPQHRAQGPGPEPTLGPLTRLEGIKVR
SEG   ......................................xxxxxxxxxxxxxxxxx.....
PRD   hhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccceeeeee
MEM   .....MMMMMMMMMMMMMMMMM......................................

SEQ   TSLPRRAPGPARLSPRGPALSPGPHAAPGAALRRSRALPLTRLLSLSGPGRLGPFWAARS
SEG   .........................xxxxxxxxxxxxxxxxxxxxxxxxxxxxx..xxxx
PRD   eeccccccccccccccccccccccccccchhhhhhcccccceeeccccccccchhhhhhc
MEM   ............................................................

SEQ   GAPAARCAPFP
SEG   xxxxxxxxx..
PRD   ccccccccccc
MEM   ...........
```

(No Prosite data available for DKFZphutel_21d15.2)

(No Pfam data available for DKFZphutel_21d15.2)

DKFZphutel_22d2
----

group: signal transduction

DKFZphutel_22d2 encodes a novel 580 amino acid putative GTP-binding protein related to the ras protein. Additionally, the putative protein contains an EF-hand for calcium-binding.

G-proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to GTP-binding proteins

complete cDNA, complete cds, potential start at Bp 64, EST hits
complete cds according to K08F11.5 and YAL048c

Sequenced by BMFZ

Locus: /map="17"

Insert length: 3247 bp
Poly A stretch at pos. 3230, no polyadenylation signal found

```
   1 CTCCTGGTGA GAGGAGTCCA CTCCGTGCGT GCGGGCGGAG GCCGGCCCCC
  51 GAGAGCCGCC GACATGAAGA AAGACGTGCG GATCCTGCTG GTGGGAGAAC
 101 CTAGAGTTGG GAAGACATCA CTGATTATGT CTCTGGTCAG TGAAGAATTT
 151 CCAGAAGAGG TTCCTCCCCG GGCAGAAGAA ATCACCATTC CAGCTGATGT
 201 CACCCCAGAG AGAGTTCCAA CACACATTGT AGATTACTCA GAAGCAGAAC
 251 AGAGTGATGA ACAACTTCAT CAAGAAATAT CTCAGGCTAA TGTCATCTGT
 301 ATAGTGTATG CCGTTAACAA CAAGCATTCT ATTGATAAGG TAACAAGTCG
 351 ATGGATTCCT CTCATAAATG AAAGAACAGA CAAAGACAGC AGGCTGCCTT
 401 TAATATTGGT TGGGAACAAA TCTGATCTGG TGGAATATAG TAGTATGGAG
 451 ACCATCCTTC CTATTATGAA CCAGTATACA GAAATAGAAA CCTGTGTGGA
 501 GTGTTCAGCG AAAAACCTGA AGAACATATC AGAGCTCTTT TATTACGCAC
 551 AGAAAGCTGT TCTTCATCCT ACAGGGCCCC TGTACTGCCC AGAGGAGAAG
 601 GAGATGAAAC CAGCTTGTAT AAAAGCCCTT ACTCGTATAT TTAAAATATC
 651 TGATCAAGAT AATGATGGTA CTCTCAATGA TGCTGAACTC AACTTCTTTC
 701 AGAGGATTTG TTTCAACACT CCATTAGCTC CTCAAGCTCT GGAGGATGTC
 751 AAGAATGTAG TCAGAAAACA TATAAGTGAT GGTGTGGCTG ACAGTGGGTT
 801 GACCCTGAAA GGTTTTCTCT TTTTACACAC ACTTTTTATC CAGAGAGGGA
 851 GACACGAAAC TACTTGGACT GTGCTTCGAC GATTTGGTTA TGATGATGAC
 901 CTGGATTTGA CACCTGAATA TTTGTTCCCC CTGCTGAAAA TACCTCCTGA
 951 TTGCACTACT GAATTAAATC ATCATGCATA TTTATTTCTC CAAAGCACCT
1001 TTGACAAGCA TGATTTGGAT AGAGACTGTG CTTTGTCACC TGATGAGCTT
1051 AAAGATTTAT TTAAAGTTTT CCCTTACATA CCTTGGGGGC CAGATGTGAA
1101 TAACACAGTT TGTACCAATG AAAGAGGCTG GATAACCTAC CAGGGATTCC
1151 TTTCCCAGTG GACGCTCACG ACTTATTTAG ATGTACAGCG GTGCCTGGAA
1201 TATTTGGGCT ATCTAGGCTA TTCAATATTG ACTGAGCAAG AGTCTCAAGC
1251 TTCAGCTGTT ACAGTGACAA GAGATAAAAA GATAGACCTG CAGAAAAAAC
1301 AAACTCAAAG AAATGTGTTC AGATGTAATG TAATTGGAGT GAAAAACTGT
1351 GGGAAAAGTG GAGTTCTTCA GGCTCTTCTT GGAAGAAACT TAATGAGGCA
1401 GAAGAAAATT CGTGAAGATC ATAAATCCTA CTATGCGATT AACACTGTTT
1451 ATGTATATGG ACAAGAGAAA TACTTGTTGT TGCATGATAT CTCAGAATCG
1501 GAATTTCTAA CTGAAGCTGA AATCATTTGT GATGTTGTAT GCCTGGTATA
1551 TGATGTCAGC AATCCCAAAT CCTTTGAATA CTGTGCCAGG ATTTTTAAGC
1601 AACACTTTAT GGACAGCAGA ATACCTTGCT TAATCGTAGC TGCAAAGTCA
1651 GACCTGCATG AAGTTAAACA AGAATACAGT ATTTCACCTA CTGATTTCTG
1701 CAGGAAACAC AAAATGCCTC CACCACAAGC CTTCACTTGC AATACTGCTG
1751 ATGCCCCCAG TAAGGATATC TTTGTTAAAT TGACAACAAT GGCCATGTAT
1801 CCGTAAGTAC TTGCTGTCTT CATTTTCATG TTGCATGGTT CATAACATTG
1851 CATGCCATTA TTAGCCATGA AGGGAATATC TTTGTCACAT AGGAATTGTT
1901 CAGCAACAGA AAGATACTTT GTAATGAGAA GGTACAAATT TGAGTAAATG
1951 CAAGTTTGGT TTGAATGCCA TAATAAAATG ATATAAACAG TGCTTCTGAC
2001 AATATCTGTA TATTTTTGAG CAGGCTGTAA CTATCTTAAT AGAATAGTAC
2051 AATAAAACAC AACCCCCCAC CCAGCATTAA AAAATAGTTT TACTGGAATA
2101 AAATGGGTTT GGCATCATGT TGTTTTATGC TTATAAAGCA TTTTCATATG
2151 AACAGAAAGT TTATATTTTT CTGTTTTTGA CCTTAGGTAT ATGAAGTTTT
2201 CTAAAATATT TTATTAATTT ATGTTGAAAT TGTGGGTATG CTTCAGTTAG
2251 GATATGTCTT TTTTAAGTGC TGTAAAGAGT AGTTGTAATT GGAATTTCTA
2301 CTGTATAAAT GTTTTACATT AAGTGTTACG AGCCACAAAT TTCATGTACA
2351 TTTATTATAT ATCTATACAT GCATATGCAC AAGCACATAA CTGTGGTCAT
2401 CTCTGTAGTT TACTAACTGC CTTAAAATTG CATGGTTCTT AATGGCATTC
2451 GCCTCAAGTA GTGTGTTTGT ATAAATTCTG TTTTGTAACA AAATAGTTTT
2501 TCAGGCAGTG CGTTTCTCAG GACTTTATAG CTTATTCTAC TTATTCTTAT
2551 GTTAGTCTCT AAATTATTTT TCTTCTTATG AAAACTACAG TGTAACACAG
```

```
2601 AGTAATAATC AAACATTGCT ATAAACCAAG AATGACATTT TTCAAAAAGG
2651 TGTTGATTTG TACAGATTTT TAAAGTCAGT TAACTTTACT GCTATTTTAT
2701 TACCTAATAC TTTTTTTAGA TGCAACAAAC CCTTGAATTT CTATTTGTAT
2751 TCGAAGACAA GTCATTCCTA TTATTATAGA ATAACCAAAA CCTTATTTAT
2801 GTTTTACCTT TGCTTTAAAA CTCTCATGTA TGTTATCTAC AGAGAGGATC
2851 ATTACAGAGA CAGACTCTCC CGAGACATGG GCCACACTGA TAGAATAGAG
2901 AATTTGAGAA AAATCTGGGT CTTTCTAAAA ACTGCTTTGT AAGTTACTTT
2951 TTCTTTATGA CTTCTGTGGG ATTTTGTTGA TATTTTCTTA GAGAATGACC
3001 AAATCTCCTT TCTTGCCATA ATTAACATTT AGTAATTATG TAGAAACGCA
3051 CTGCTTGGTC AGGCTTCCTG CCTAGCTATA TATTACGTTG TCTTCCTTAC
3101 TACATAAATG TACTTCTTTA ATCTTGTGAT TACAGTAACT GCAAGTGTGT
3151 TTTTACATCT GCATTTTTAA AACATTTTAC TGTAATTCTG TTGTGTGTGT
3201 GTGTGTTATA TGATAAATGT ACATACATGG AAAAAAAAAA AAAAAAA
```

BLAST Results

Entry AC004527 from database EMBL:
*** SEQUENCING IN PROGRESS *** NF1-related locus, Direct Submission;
HTGS phase 1, 10 unordered pieces.
Score = 1899, P = 1.1e-78, identities = 387/396

Entry HS148355 from database EMBL:
human STS SHGC-31220.
Score = 1826, P = 7.5e-78, identities = 388/406

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 64 bp to 1803 bp; peptide length: 580
Category: similarity to known protein

```
  1 MKKDVRILLV GEPRVGKTSL IMSLVSEEFP EEVPPRAEEI TIPADVTPER
 51 VPTHIVDYSE AEQSDEQLHQ EISQANVICI VYAVNNKHSI DKVTSRWIPL
101 INERTDKDSR LPLILVGNKS DLVEYSSMET ILPIMNQYTE IETCVECSAK
151 NLKNISELFY YAQKAVLHPT GPLYCPEEKE MKPACIKALT RIFKISDQDN
201 DGTLNDAELN FFQRICFNTP LAPQALEDVK NVVRKHISDG VADSGLTLKG
251 FLFLHTLFIQ RGRHETTWTV LRRFGYDDDL DLTPEYLFPL LKIPPDCTTE
301 LNHHAYLFLQ STFDKHDLDR DCALSPDELK DLFKVFPYIP WGPDVNNTVC
351 TNERGWITYQ GFLSQWTLTT YLDVQRCLEY LGYLGYSILT EQESQASAVT
401 VTRDKKIDLQ KKQTQRNVFR CNVIGVKNCG KSGVLQALLG RNLMRQKKIR
451 EDHKSYYAIN TVYVYGQEKY LLLHDISESE FLTEAEIICD VVCLVYDVSN
501 PKSFEYCARI FKQHFMDSRI PCLIVAAKSD LHEVKQEYSI SPTDFCRKHK
551 MPPPQAFTCN TADAPSKDIF VKLTTMAMYP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphute1_22d2, frame 1

TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid K08F11., N = 1, Score = 1357, P = 1.1e-138

TREMBL:SPCC320_4 gene: "SPCC320.04c"; product: "hypothetical protein"; S.pombe chromosome III cosmid c320., N = 1, Score = 889, P = 4.4e-89

TREMBL:CEUC47C12_3 gene: "C47C12.4"; Caenorhabditis elegans cosmid C47C12., N = 2, Score = 408, P = 5.6e-74

PIR:S51971 probable membrane protein YAL048c - yeast (Saccharomyces cerevisiae), N = 1, Score = 677, P = 1.3e-66

>TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid
    K08F11.
        Length = 625

HSPs:

```
 Score = 1357 (203.6 bits), Expect = 1.1e-138, P = 1.1e-138
 Identities = 263/582 (45%), Positives = 380/582 (65%)

Query:     4 DVRILLVGEPRVGKTSLIMSLVSEEFPEEVPPRAEEITIPADVTPERVPTHIVDYSEAEQ 63
             DVRI+L+G+   GKTSL+MSL+ +E+ + VP R + + IPADVTPE V T IVD S  E+
Sbjct:     9 DVRIVLIGDEGCGKTSLVMSLLEDEWVDAVPRRLDRVLIPADVTPENVTTSIVDLSIKEE 68

Query:    64 SDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLINERTDKDSRLPLILVGNKSDLV 123
              +  +  EI QANVIC+VY+V ++ ++D + ++W+PLI +    +   P+ILVGNKSD
Sbjct:    69 DENWIVSEIRQANVICVVYSVTDESTVDGIQTKWLPLIRQSFGEYHETPVILVGNKSDGT 128

Query:   124 EYSSMETILPIMNQYTEIETCVECSAKNLKNISELFYYAQKAVLHPTGPLYCPEEKEMKP 183
               ++ + ILPIM   TE+ETCVECSA+ +KN+SE+FYYAQKAV++PT PLY  + K++
Sbjct:   129 A-NNTDKILPIMEANTEVETCVECSARTMKNVSEIFYYAQKAVIYPTRPLYDADTKQLTD 187

Query:   184 ACIKALTRIFKISDQDNDGTLNDAELNFFQRICFNTPLAPQALEDVKNVVRKHISDGVAD 243
                KAL R+FKI D+DNDG L+D ELN FQ++CF  PL   ALEDVK  V      DGVA+
Sbjct:   188 RARKALIRVFKICDRDNDGYLSDTELNDFQKLCFGIPLTSTALEDVKRAVSDGCPDGVAN 247

Query:   244 SGLTLKGFLFLHTLFIQRGRHETTWTVLRRFGYDDDLDLTPEYLFPLLKIPPDCTTELNH 303
               L L GFL+LH LFI+RGRHETTW VLR+FGY+  L L+ +YL+P + IP  C+TEL+
Sbjct:   248 DSLMLAGFLYLHLLFIERGRHETTWAVLRKFGYETSLKLSEDYLYPRITIPVGCSTELSP 307

Query:   304 HAYLFLQSTFDKHDLDRDCALSPDELKDLFKVFPYIPWGPDVNNTVCTNERGWITYQGFL 363
                 F+ + F+K+D D+D  LSP EL++LF V P       D    + TN+RGW+TY G++
Sbjct:   308 EGVQFVSALFEKYDEDKDGCLSPSELQNLFSVCPVPVITKDNILALETNQRGWLTYNGYM 367

Query:   364 SQWTLTTYLDVQRCLEYLGYLGYSILTEQESQAS----AVTVTRDKKIDLQKKQTQRNVF 419
             + W +TT +++ +  E L YLG+ +        +A     ++ VTR++K DL+    T R VF
Sbjct:   368 AYWNMTTLINLTQTFEQLAYLGFPVGRSGPGRAGNTLDSIRVTRERKKDLENHGTDRKVF 427

Query:   420 RCNVIGVKNCGKSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLLHDI--- 476
             +C V+G K+ GK+  +Q+L GR +    +I   H S + IN V V  + KYLLL ++
Sbjct:   428 QCLVVGAKDAGKTVFMQSLAGRGMADVAQIGRRH-SPFVINRVRVKEESKYLLLREVDVL 486

Query:   477 SESEFLTEAEIICDVVCLVYDVSNPKSFEYCARIFKQHFMDSRIPCLIVAAKSDLHEVKQ 536
             S  + L   E   DVV  +YD+SNP SF +CA +++++F  ++ PC+++A K +  EV Q
Sbjct:   487 SPQDALGSGETSADVVAFLYDISNPDSFAFCATVYQKYFYRTKTPCVMIATKVEREEVDQ 546

Query:   537 EYSISPTDFCRKHKMPPPQAFTCNTADAPSKDIFVKLTTMAMYP 580
              + + P +FCR+ ++P P  F+        S  IF +L  MA+YP
Sbjct:   547 RWEVPPEEFCRQFELPKPIKFSTGNIGQSSSPIFEQLAMMAVYP 590
```

Pedant information for DKFZphute1_22d2, frame 1

Report for DKFZphute1_22d2.1

```
[LENGTH]        580
[MW]            66541.61
[pI]            5.56
[HOMOL]         TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid K08F11. 1e-149
[FUNCAT]        99 unclassified proteins      [S. cerevisiae, YAL048c] 5e-81
[FUNCAT]        03.04 budding, cell polarity and filament formation [S. cerevisiae, YKR055w] 3e-11
[FUNCAT]        03.99 other cell growth, cell division and dna synthesis activities      [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]        10.04.07 g-proteins    [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]        03.10 sporulation and germination    [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]        11.01 stress response [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]        03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]        01.03.13 regulation of nucleotide metabolism        [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]        01.05.04 regulation of carbohydrate utilization      [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]        30.03 organization of cytoplasm       [S. cerevisiae, YOR101w] 4e-08
[FUNCAT]        11.10 cell death         [S. cerevisiae, YOR101w] 4e-08
[FUNCAT]        10.02.07 g-proteins    [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]        30.04 organization of cytoskeleton    [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]        30.08 organization of golgi    [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]        08.07 vesicular transport (golgi network, etc.)      [S. cerevisiae, YFL005w] 9e-08
[FUNCAT]        30.09 organization of intracellular transport vesicles      [S. cerevisiae, YFL005w] 9e-08
[FUNCAT]        30.02 organization of plasma membrane        [S. cerevisiae, YFL005w] 9e-08
[FUNCAT]        08.13 vacuolar transport      [S. cerevisiae, YNL093w] 1e-07
```

[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YNL093w] 1e-07
[FUNCAT] 08.19 cellular import [S. cerevisiae, YNL093w] 1e-07
[FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YLR229c] 8e-07
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YLR229c] 8e-07
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 3e-06
[FUNCAT] 09.09 biogenesis of intracellular transport vesicles [S. cerevisiae, YGL210w] 9e-04
[BLOCKS] BL00410A Dynamin family proteins
[SCOP] d1p1k__ 3.25.1.3.1 cH-p21 Ras protein [human (Homo sapiens) 2e-42
[SCOP] d1guaa_ 3.25.1.3.10 Rap1A (Human (Homo sapiens) 5e-59
[PIRKW] transmembrane protein 1e-79
[PIRKW] membrane trafficking 2e-06
[PIRKW] acetylated amino end 3e-09
[PIRKW] prenylated cysteine 3e-09
[PIRKW] signal transduction 1e-07
[PIRKW] transforming protein 3e-09
[PIRKW] immediate-early protein 8e-06
[PIRKW] alternative splicing 4e-08
[PIRKW] P-loop 1e-10
[PIRKW] lipoprotein 7e-10
[PIRKW] proto-oncogene 3e-09
[PIRKW] methylated carboxyl end 3e-09
[PIRKW] membrane protein 3e-09
[PIRKW] GTP binding 1e-10
[PIRKW] thiolester bond 7e-10
[SUPFAM] ras transforming protein 1e-10
[PROSITE] ATP_GTP_A 2
[PROSITE] MYRISTYL 3
[PROSITE] EF_HAND 1
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 14
[PROSITE] TYR_PHOSPHO_SITE 4
[PROSITE] PKC_PHOSPHO_SITE 5
[PROSITE] ASN_GLYCOSYLATION 3
[PFAM] Ras family (contains ATP/GTP binding P-loop)
[KW] Irregular
[KW] 3D

```
SEQ    MKKDVRILLVGEPRVGKTSLIMSLVSEEFPEEVPPRAEEITIPADVTPERVPTHIVDYSE
1jai-  ...EEEEEEEETTTTCHHHHHHHHHHCCCCCCCCCCCEEEEEEEETTEEEEEEEEECCC

SEQ    AEQSDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLINERTDKDSRLPLILVGNKS
1jai-  CGGGHHHHHHHHHHTTEEEEEEETTTHHHHHHH-HHHHHHHHHHHCTTT-TCEEEEEETT

SEQ    DLVEYSSMETILPIMNQYTEIETCVECSAKNLKNISELFYYAQKAVLHPTGPLYCPEEKE
1jai-  TTTTTTTTHHHHHHHHHHHCCCE-EECTTTTTTTHHHHHH....................

SEQ    MKPACIKALTRIFKISDQDNDGTLNDAELNFFQRICFNTPLAPQALEDVKNVVRKHISDG
1jai-  ............................................................

SEQ    VADSGLTLKGFLFLHTLFIQRGRHETTWTVLRRFGYDDDLDLTPEYLFPLLKIPPDCTTE
1jai-  ............................................................

SEQ    LNHHAYLFLQSTFDKHDLDRDCALSPDELKDLFKVFPYIPWGPDVNNTVCTNERGWITYQ
1jai-  ............................................................

SEQ    GFLSQWTLTTYLDVQRCLEYLGYLGYSILTEQESQASAVTVTRDKKIDLQKKQTQRNVFR
1jai-  ............................................................

SEQ    CNVIGVKNCGKSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLLHDISESE
1jai-  ............................................................

SEQ    FLTEAEIICDVVCLVYDVSNPKSFEYCARIFKQHFMDSRIPCLIVAAKSDLHEVKQEYSI
1jai-  ............................................................

SEQ    SPTDFCRKHKMPPPQAFTCNTADAPSKDIFVKLTTMAMYP
1jai-  ........................................
```

Prosite for DKFZphute1_22d2.1

PS00001	118->122	ASN_GLYCOSYLATION	PDOC00001
PS00001	154->158	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00004	411->415	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005

PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	311->315	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00007	153->161	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->384	TYR_PHOSPHO_SITE	PDOC00007
PS00007	153->162	TYR_PHOSPHO_SITE	PDOC00007
PS00007	448->457	TYR_PHOSPHO_SITE	PDOC00007
PS00008	240->246	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008
PS00008	433->439	MYRISTYL	PDOC00008
PS00017	11->19	ATP_GTP_A	PDOC00017
PS00017	425->433	ATP_GTP_A	PDOC00017
PS00018	197->210	EF_HAND	PDOC00018

Pfam for DKFZphutel_22d2.1

```
HMM_NAME     Ras family (contains ATP/GTP binding P-loop)

HMM              *KLVLIGDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGKtIK
                   ++L+G+  VGK++L   ++  EF+EE +P   ++  T   ++  +++
Query          6  RILLVGEPRVGKTSLIMSLVSEEFPEE-VPPR-AEEITIPADVTPERVP    52

HMM              LQIWDTAGQERYRsMRPMYYRGAMGFMLVYDITNRqSFENIr.NWweEIr
                   I D    E+ +    + + +A+++ +VY+++N+ S ++++ +W++ I+
Query         53 THIVDYSEAEQSDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLIN   102

HMM              RHCDrDENVPIMLVGNKCDLEDQRQVStEEGQeFAREWGAIPFMETSAKT
                 +  D+D+  P +LVGNK+DL +   ++T         +     +E+SAK+
Query        103 ERTDKDSRLPLILVGNKSDLVEYSSMETILPIMNQYTEI-ETCVECSAKN   151

HMM              NiNVEEAFMEIvReIlqrMqeqNqteNinidQpsrnrkrCCCIM*
                   N+ E F+ + +++L +       +++  +++++ +   C+
Query        152 LKNISELFYYAQKAVLHPT------GPLYCPEEKEMK-PACI--   186
```

DKFZphutel_22e12

group: signal transduction

DKFZphutel_22e12 encodes a novel 92 amino acid protein, with similarity to yeast, C.elegans, Drosophila and mammalian proteins.

The Drosophila cni and mammalian cornicon proteins are part of a signal transduction pathway involving hte EGF-receptor.

The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

strong similarity to S.cerevisiae YGL054c and cornichon

complete cDNA, complete cds, EST hits
cornicon is requiered for signal transduction in the EGF-receptor signal processing

Sequenced by BMFZ

Locus: unknown

Insert length: 519 bp
Poly A stretch at pos. 499, no polyadenylation signal found

```
  1 GTCGGGGCAT CCGAGCGGGT TTGACGGAAG GAGCGGCGGC GACGGAGGAG
 51 GAGGATGGAG GCGGTGGTGT TCGTCTTCTC TCTCCTCGAT TGTTGCGCGC
101 TCATCTTCCT CTCGGTCTAC TTCATAATTA CATTGTCTGA TTTAGAATGT
151 GATTACATTA ATGCTAGATC ATGTTGCTCA AAATTAAACA AGTGGGTAAT
201 TCCAGAATTG ATTGGCCATA CCATTGTCAC TGTATTACTG CTCATGTCAT
251 TGCACTGGTT CATCTTCCTT CTCAACTTAC CTGTTGCCAC TTGGAATATA
301 TATCGTATGA TCTTAGCTTT GATAAATGAC TGAAGCTGGA GAAGCCGTGG
351 TTGAAGTCAG CCTACACTAC AGTGCACAGT TGAGGAGCCA GAGACTTCTT
401 AAATCATCCT TAGAACCGTG ACCATAGCAG TATATATTTT CCTCTTGGAA
451 CAAAAAACTA TTTTTGCTGT ATTTTTACCA TATAAAGTAT TTAAAAAACA
501 TGAAAAAAAA AAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

95300228:
cornichon and the EGF receptor signaling process are necessary for both anterior-posterior
and dorsal-ventral pattern formation in Drosophila.

Peptide information for frame 1

ORF from 55 bp to 330 bp; peptide length: 92
Category: strong similarity to known protein

```
 1 MEAVVFVFSL LDCCALIFLS VYFIITLSDL ECDYINARSC CSKLNKWVIP
51 ELIGHTIVTV LLLMSLHWFI FLLNLPVATW NIYRMILALI ND
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_22e12, frame 1

PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae), N = 2, Score = 185, P = 5.7e-17

TREMBL:SPAC2C4_5 gene: "SPAC2C4.05"; product: "cornichon homolog";

S.pombe chromosome I cosmid c2C4., N = 1, Score = 163, P = 3.7e-12

PIR:S46084 probable membrane protein YBR210w - yeast (Saccharomyces cerevisiae), N = 1, Score = 162, P = 4.8e-12

TREMBL:AF104398_1 product: "cornichon"; Homo sapiens cornichon mRNA, complete cds., N = 1, Score = 141, P = 8e-10

SWISSPROT:CNI_DROVI CORNICHON PROTEIN., N = 1, Score = 139, P = 1.3e-09

```
>PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces
     cerevisiae)
          Length = 138

  HSPs:

 Score = 185 (27.8 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17
 Identities = 35/85 (41%), Positives = 56/85 (65%)

Query:     1 MEAVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCCSKLNKWVIPELIGHTIVTV 60
             M A +F+ +++  C  +F  V+F I  +DLE DYIN    CSK+NK + PE   H +++
Sbjct:     1 MGAWLFILAVVVNCINLFGQVHFTILYADLEADYINPIELCSKVNKLITPEAALHGALSL 60

Query:    61 LLLMSLHWFIFLLNLPVATWNIYRM 85
             L L++ +WF+FLLNLPV  +N+ ++
Sbjct:    61 LFLLNGYWFVFLLNLPVLAYNLNKI 85

 Score = 37 (5.6 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17
 Identities = 7/9 (77%), Positives = 9/9 (100%)

Query:    82 IYRMILALI 90
             +YRMI+ALI
Sbjct:   123 LYRMIMALI 131
```

Pedant information for DKFZphutel_22e12, frame 1

Report for DKFZphutel_22e12.1

```
[LENGTH]        92
[MW]            10614.98
[pI]            5.04
[HOMOL]         PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)
5e-14
[FUNCAT]        03.04 budding, cell polarity and filament formation [S. cerevisiae, YGL054c]
2e-15
[PIRKW]         transmembrane protein 2e-11
[PROSITE]       CK2_PHOSPHO_SITE       3
[KW]            SIGNAL_PEPTIDE 33
[KW]            TRANSMEMBRANE  2

SEQ     MEAVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCCSKLNKWVIPELIGHTIVTV
PRD     ccchhhhhhhhhhhhhhhhhhhheeeccccccccccccccccccceeehhhhhhhhhhh
MEM     ...............................................MMMMMMMMMM

SEQ     LLLMSLHWFIFLLNLPVATWNIYRMILALIND
PRD     hhhhhhhheeecccccchhhhhhhhhhhhccc
MEM     MMMMMMMMMMMMMMMMMMMM..MMMMMMM....
```

Prosite for DKFZphutel_22e12.1

PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	26->30	CK2_PHOSPHO_SITE	PDOC00006
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphutel_22e12.1)

DKFZphutel_22n2
----

group: uterus derived

DKFZphutel_22n2 encodes a novel 304 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="553.3 cR from top of Chr11 linkage group"

Insert length: 1556 bp
Poly A stretch at pos. 1534, no polyadenylation signal found

```
   1 ACAACAGGCT GGTTGCTTGG CGTGGAATCC TAAAGTGGCC TGGCTTTGAG
  51 ACTGGAGTGA GACCCCAGCC CTAGGCTGGG GTTCTTTCCA TTATAGAGGA
 101 GACGGATTCA GAAGGGCTAC AGACCAAGGT TGTTGAAAAC CAGACATATG
 151 ATGAGCGTCT AGAGATTAAC GACTCCGAAG AGGTTGCAAG TATTTATACT
 201 CCAACCCCAA GACACCAAGG ACTTCCTCGT TCTGCCCATC TTCCTAACAA
 251 GGCTATGGCT GATAACAGCA GTGATGAGTG TGAAGAGGAA AATAACAAGG
 301 AGAAGAAGAA GACCTCACAG TTGACACCTC AACGGGGCTT TAGTGAAAAT
 351 GAGGATGACG ATGATGATGA TGATGATTCA TCTGAAACTG ATTCTGATTC
 401 TGATGATGAT GATGAAGAGC ATGGAGCCCC TCTGGAAGGG GCCTATGACC
 451 CTGCAGACTA TGAGCATTTG CCAGTTTCTG CTGAAATTAA GGAACTCTTC
 501 CAGTACATCA GTAGGTACAC ACCTCAGTTG ATTGACCTGG ACCACAAACT
 551 GAAGCCTTTC ATTCCTGATT TTATCCCAGC TGTCGGGGAT ATTGATGCAT
 601 TCTTAAAGGT CCCACGTCCT GATGGAAAGC CTGACAACCT TGGCCTATTG
 651 GTATTGGATG AACCTTCTAC AAAGCAGTCA GACCCTACGG TGCTCTCACT
 701 CTGGTTAACA GAGAATTCTA AGCAGCACAA CATCACACAA CATATGAAAG
 751 TAAAAAGCCT AGAAGATGCA GAAAAGAATC CCAAAGCCAT TGACACGTGG
 801 ATTGAGAGCA TCTCTGAATT ACACCGTTCT AAGCCCCCTG CGACTGTGCA
 851 CTACACCAGG CCCATGCCCG ACATTGACAC GCTGATGCAG GAATGGTCCC
 901 CGGAGTTTGA AGAGCTTTTG GGCAAGGTAA GCCTGCCCAC GGCAGAGATT
 951 GATTGCAGCC TGGCAGAGTA CATTGACATG ATCTGTGCCA TTCTAGACAT
1001 CCCTGTCTAC AAGAGTCGGA TCCAGTCCCT CCATCTGCTC TTTTCCCTCT
1051 ACTCAGAATT CAAGAACTCA CAGCATTTTA AAGCTCTCGC TGAAGGCAAG
1101 AAAGCATTCA CTCCTTCATC CAATTCCACC TCCCAAGCTG GAGACATGGA
1151 GACATTAACC TTCAGCTGAG ACACTTCCCA AGCTGCTGTT TCAAGGCTGA
1201 GCTGGCCCCT CTGCCCCAGC TGAGATGGAC AGATCGTTGT CAGCTACTTG
1251 ATGTCCTTGC CCATGCCACA GCTTGGCTCA GGGGCAGTGC ATGTCCTGCT
1301 GCCCTCTCTG CCAGAGGGCA CAGAACATGT TTGTTTAATG AACCTGCCTG
1351 CCTCAGATTG CTGTCCCCGG GGAGTTAATG CATCTACACC ACTGTGGGGA
1401 TTTGAGTTAT AAGAATTGGA ATTTCTGAGA TCCCATGGAG GTTAGATTGG
1451 GAGGAAAGCT TAAAAGATGT CCTTTTTGTG AGAGGGATGG AATTGTTTTC
1501 TTTCATTCGT AAAGTTAGTG AGTAAAGATT TTATAAATCA AAAAAAAAAA
1551 AAAAAA
```

BLAST Results

Entry HS188252 from database EMBL:
human STS WI-12265.
Score = 2554, P = 4.1e-109, identities = 556/587

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 255 bp to 1166 bp; peptide length: 304
Category: putative protein

```
  1 MADNSSDECE EENNKEKKKT SQLTPQRGFS ENEDDDDDDD DSSETDSDSD
 51 DDDEEHGAPL EGAYDPADYE HLPVSAEIKE LFQYISRYTP QLIDLDHKLK
101 PFIPDFIPAV GDIDAFLKVP RPDGKPDNLG LLVLDEPSTK QSDPTVLSLW
151 LTENSKQHNI TQHMKVKSLE DAEKNPKAID TWIESISELH RSKPPATVHY
201 TRPMPDIDTL MQEWSPEFEE LLGKVSLPTA EIDCSLAEYI DMICAILDIP
251 VYKSRIQSLH LLFSLYSEFK NSQHFKALAE GKKAFTPSSN STSQAGDMET
301 LTFS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_22n2, frame 3

PIR:S38149 SIS2 protein - yeast (Saccharomyces cerevisiae), N = 1,
Score = 132, P = 1e-05

```
>PIR:S38149 SIS2 protein - yeast (Saccharomyces cerevisiae)
          Length = 562

  HSPs:

 Score = 132 (19.8 bits), Expect = 1.0e-05, P = 1.0e-05
 Identities = 24/63 (38%), Positives = 35/63 (55%)

Query:      3 DNSSDECEEENNKEKKKTSQLTPQRGFSENEDDDDDDDDSSETDSDSDDDDEEHGAPLEG 62
              +    DE EEE++ E++ T              +++DDDDDDD  + D D DDD++E  A   G
Sbjct:    497 EEDDDEDEEEDDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDDDDDEDEDEAETPG 556

Query:     63 AYD 65
                D
Sbjct:    557 IID 559

 Score = 122 (18.3 bits), Expect = 1.4e-04, P = 1.4e-04
 Identities = 20/52 (38%), Positives = 33/52 (63%)

Query:      4 NSSDECEEENNKEKKKTSQLTPQRGFSENEDDDDDDDDSSETDSDSDDDDEE 55
              N+ +E ++E+ +E      + T  +  + N+DDDDDDD  + D D DDD++
Sbjct:    494 NNEEEDDDEDEEEDDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDDDDD 545
```

Pedant information for DKFZphutel_22n2, frame 3

Report for DKFZphutel_22n2.3

```
[LENGTH]      304
[MW]          34285.85
[pI]          4.37
[PROSITE]     AMIDATION      1
[PROSITE]     CAMP_PHOSPHO_SITE      2
[PROSITE]     CK2_PHOSPHO_SITE       10
[PROSITE]     PKC_PHOSPHO_SITE       1
[PROSITE]     ASN_GLYCOSYLATION      3
[KW]          All_Alpha
[KW]          LOW_COMPLEXITY    11.84 %

SEQ   MADNSSDECEEENNKEKKKTSQLTPQRGFSENEDDDDDDDDSSETDSDSDDDDEEHGAPL
SEG   .........xxxxxxxxxx..........xxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD   cccccccchhhhhhchhhhhhcccccccccccccccccccccccccccccccccccccccc

SEQ   EGAYDPADYEHLPVSAEIKELFQYISRYTPQLIDLDHKLKPFIPDFIPAVGDIDAFLKVP
SEG   ............................................................
PRD   cccccccccccchhhhhhhhhhhhhhhcccccccccccccccccccccccccccceeecc

SEQ   RPDGKPDNLGLLVLDEPSTKQSDPTVLSLWLTENSKQHNITQHMKVKSLEDAEKNPKAID
SEG   ............................................................
PRD   cccccccceeeeeccccccccccccchhhhhhcccccccccccchhhhhhhhhcccccch

SEQ   TWIESISELHRSKPPATVHYTRPMPDIDTLMQEWSPEFEELLGKVSLPTAEIDCSLAEYI
SEG   ............................................................
PRD   hhhhhhhhhhccccceeeeecccccchhhhhhccchhhhhhccccccccccchhhhhhh

SEQ   DMICAILDIPVYKSRIQSLHLLFSLYSEFKNSQHFKALAEGKKAFTPSSNSTSQAGDMET
SEG   ............................................................
```

PRD hhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhcccccccccccccccccccc

SEQ LTFS
SEG ....
PRD cccc

Prosite for DKFZphutel_22n2.3

PS00001	4->8	ASN_GLYCOSYLATION	PDOC00001
PS00001	159->163	ASN_GLYCOSYLATION	PDOC00001
PS00001	290->294	ASN_GLYCOSYLATION	PDOC00001
PS00004	17->21	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	18->22	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDOC00006
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	43->47	CK2_PHOSPHO_SITE	PDOC00006
PS00006	45->49	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	168->172	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	185->189	CK2_PHOSPHO_SITE	PDOC00006
PS00006	235->239	CK2_PHOSPHO_SITE	PDOC00006
PS00009	280->284	AMIDATION	PDOC00009

(No Pfam data available for DKFZphutel_22n2.3)

DKFZphutel_22o2
----

group: uterus derived

DKFZphutel_22o2 encodes a novel 537 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to S.pombe SPBC3E7.03c

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: map="11p15.5"

Insert length: 2714 bp
Poly A stretch at pos. 2695, polyadenylation signal at pos. 2677

```
   1 GCAGGGCACG GTGGGGGCTG AGATCGTTTC CTGTTGGAAC TTCTGGCCCA
  51 AGAAGCGCGG GTCACAAGGA GAGGGGTCAG TTCGGTTCAG AGCGACTCAG
 101 CCCCTCGACT CGGGTCTTAA AACCTCCGAG CCGCCAGTTC TGCCTCAGGC
 151 CGCGCCCCCT TAAAGCGCCA CCAGACGCTG CGCCCCGTTA AAGCGCCACC
 201 AGACGCCGCG CCCCGTCCCG GCCTCCCCCG CGCGCTGGCG CGGGGCTTTC
 251 TGGGCCAGGG CGGGGCCGGC GAACTGCGGC CCGGAACGGC TGAGGAAGGG
 301 CCCGTCCCGC CTTCCCCGGC GCGCCATGGA GCCCCGGGCG GTTGCAGAAG
 351 CCGTGGAGAC GGGTGAGGAG GATGTGATTA TGGAAGCTCT GCGGTCATAC
 401 AACCAGGAGC ACTCCCAGAG CTTCACGTTT GATGATGCCC AACAGGAGGA
 451 CCGGAAGAGA CTGGCGGAGC TGCTGGTCTC CGTCCTGGAA CAGGGCTTGC
 501 CACCCTCCCA CCGTGTCATC TGGCTGCAGA GTGTCCGAAT CCTGTCCCGG
 551 GACCGCAACT GCCTGGACCC GTTCACCAGC CGCCAGAGCC TGCAGGCACT
 601 AGCCTGCTAT GCTGACATCT CTGTCTCTGA GGGGTCCGTC CCAGAGTCCG
 651 CAGACATGGA TGTTGTACTG GAGTCCCTCA AGTGCCTGTG CAACCTCGTG
 701 CTCAGCAGCC CTGTGGCACA GATGCTGGCA GCAGAGGCCC GCCTAGTGGT
 751 GAAGCTCACA GAGCGTGTGG GGCTGTACCG TGAGAGGAGC TTCCCCCACG
 801 ATGTCCAGTT CTTTGACTTG CGGCTCCTCT TCCTGCTAAC GGCACTCCGC
 851 ACCGATGTGC GCCAGCAGCT GTTTCAGGAG CTGAAAGGAG TGCGCCTGCT
 901 AACTGACACA CTGGAGCTGA CGCTGGGGGT GACTCCTGAA GGGAACCCCC
 951 CACCCACGCT CCTTCCTTCC CAAGAGACTG AGCGGGCCAT GGAGATCCTC
1001 AAAGTGCTCT TCAACATCAC CCTGGACTCC ATCAAGGGGG AGGTGGACGA
1051 GGAAGACGCT GCCCTTTACC GACACCTGGG GACCCTTCTC CGGCACTGTG
1101 TGATGATCGC TACTGCTGGA GACCGCACAG AGGAGTTCCA CGGCCACGCA
1151 GTGAACCTCC TGGGGAACTT GCCCCTCAAG TGTCTGGATG TTCTCCTCAC
1201 CCTGGAGCCA CATGGAGACT CCACGGAGTT CATGGGAGTG AATATGGATG
1251 TGATTCGTGC CCTCCTCATC TTCCTAGAGA AGCGTTTGCA CAAGACACAC
1301 AGGCTGAAGG AGAGTGTAGC TCCCGTGCTG AGCGTGCTGA CTGAATGTGC
1351 CCGGATGCAC CGCCCAGCCA GGAAGTTCCT GAAGGCCCAG GGATGGCCAC
1401 CTCCCCAGGT GCTGCCCCCT CTGCGGGATG TGAGGACACG GCCTGAGGTT
1451 GGGGAGATGC TGCGGAACAA GCTTGTCCGC CTCATGACAC ACCTGGACAC
1501 AGATGTGAAG AGGGTGGCTG CCGAGTTCTT GTTTGTCCTG TGCTCTGAGA
1551 GTGTGCCCCG ATTCATCAAG TACACAGGCT ATGGGAATGC TGCTGGCCTT
1601 CTGGCTGCCA GGGGCCTCAT GGCAGGAGGC CGGCCCGAGG GCCAGTACTC
1651 AGAGGATGAG GACACAGACA CAGATGAGTA CAAGGAAGCC AAAGCCAGCA
1701 TAAACCCTGT GACCGGGAGG GTGGAGGAGA AGCCGCCTAA CCCTATGGAG
1751 GGCATGACAG AGGAGCAGAA GGAGCACGAG GCCATGAAGC TGGTGACCAT
1801 GTTTGACAAG CTCTCCAGGA ACAGAGTCAT CCAGCCAATG GGGATGAGTC
1851 CCCGGGGTCA TCTTACGTCC CTGCAGGATG CCATGTGCGA GACTATGGAG
1901 CAGCAGCTCT CCTCGGACCC TGACTCGGAC CCTGACTGAG GATGGCAGCT
1951 CTTCTGCTCC CCCATCAGGA CTGGTGCTGC TTCCAGAGAC TTCCTTGGGG
2001 TTGCAACCTG GGGAAGCCAC ATCCCACTGG ATCCACACCC GCCCCCACTT
2051 CTCCATCTTA GAAACCCCTT CTCTTGACTC CCGTTCTGTT CATGATTTGC
2101 CTCTGGTCCA GTTTCTCATC TCTGGACTGC AACGGTCTTC TTGTGCTAGA
2151 ACTCAGGCTC AGCCTCGAAT TCCACAGACG AAGTACTTTC TTTTGTCTGC
2201 GCCAAGAGGA ATGTGTTCAG AAGCTGCTGC CTGAGGGCAG GGCCTACCTG
2251 GGCACACAGA AGAGCATATG GGAGGGCAGG GGTTTGGGTG TGGGTGCACA
2301 CAAAGCAAGC ACCATCTGGG ATTGGCACAC TGGCAGAGCC AGTGTGTTGG
2351 GGTATGTGCT GCACTTCCCA GGGAGAAAAC CTGTCAGAAC TTTCCATACG
2401 AGTATATCAG AACACACCCT TCCAAGGTAT GTATGCTCTG TTGTTCCTGT
2451 CCTGTCTTCA CTGAGCGCAG GGCTGGAGGC CTCTTAGACA TTCTCCTTGG
2501 TCCTCGTTCA GCTGCCCACT GTAGTATCCA CAGTGCCCGA GTTCTCGCTG
2551 GTTTTGGCAA TTAAACCTCC TTCCTACTGG TTTAGACTAC ACTTACAACA
2601 AGGAAAATGC CCCTCGTGTG ACCATAGATT GAGATTTATA CCACATACCA
2651 CACATAGCCA CAGAAACATC ATCTTGAAAT AAAGAAGAGT TTTGGACAAA
2701 AAAAAAAAAA AAAA
```

BLAST Results

Entry AF015416 from database EMBL:
Homo sapiens chromosome 11 from 11p15.5 region, complete sequence.
Score = 3356, P = 2.0e-144, identities = 672/673

Entry HS263253 from database EMBL:
human STS SHGC-15914.
Score = 1143, P = 9.0e-46, identities = 245/255

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 326 bp to 1936 bp; peptide length: 537
Category: similarity to unknown protein

```
  1 MEPRAVAEAV ETGEEDVIME ALRSYNQEHS QSFTFDDAQQ EDRKRLAELL
 51 VSVLEQGLPP SHRVIWLQSV RILSRDRNCL DPFTSRQSLQ ALACYADISV
101 SEGSVPESAD MDVVLESLKC LCNLVLSSPV AQMLAAEARL VVKLTERVGL
151 YRERSFPHDV QFFDLRLLFL LTALRTDVRQ QLFQELKGVR LLTDTLELTL
201 GVTPEGNPPP TLLPSQETER AMEILKVLFN ITLDSIKGEV DEEDAALYRH
251 LGTLLRHCVM IATAGDRTEE FHGHAVNLLG NLPLKCLDVL LTLEPHGDST
301 EFMGVNMDVI RALLIFLEKR LHKTHRLKES VAPVLSVLTE CARMHRPARK
351 FLKAQGWPPP QVLPPLRDVR TRPEVGEMLR NKLVRLMTHL DTDVKRVAAE
401 FLFVLCSESV PRFIKYTGYG NAAGLLAARG LMAGGRPEGQ YSEDEDTDTD
451 EYKEAKASIN PVTGRVEEKP PNPMEGMTEE QKEHEAMKLV TMFDKLSRNR
501 VIQPMGMSPR GHLTSLQDAM CETMEQQLSS DPDSDPD
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphute1_22o2, frame 2

TREMBL:SPBC3E7_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";
S.pombe chromosome II cosmid c3E7., N = 1, Score = 112, P = 0.0023

>TREMBL:SPBC3E7_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";
S.pombe chromosome II cosmid c3E7.
Length = 362

HSPs:

Score = 112 (16.8 bits), Expect = 2.3e-03, P = 2.3e-03
Identities = 71/289 (24%), Positives = 124/289 (42%)

```
Query:  215 SQETERAM-EILKVLFNITLDSIKGEVDEEDAALYRHLGTLLRHCVMIATAGDRTEEFHG 273
            SQ+ E  + EIL++LF I+  S   E DE+   L      L+      +   +
Sbjct:   12 SQDNEMVLTEILRLLFPISKRSYLKEEDEQKILL------LVIEIWASSLNNNPNSPLRW 65

Query:  274 HAVN-LLG-NLPLKCLDVLLTLEPHGDSTEFMGVNMDVIRALLIFLEKRLHKTH----RL 327
            HA N LL  NL L  LD  + +      T    +   +I   + +LEK L+       +
Sbjct:   66 HATNALLSFNLQLLSLDQAIYVSEIACQT----LQSILISREVEYLEKGLNLCFDIAAKY 121

Query:  328 KESVAPVLSVLTECARMHRPARKFLKAQGWPPPQVLPPLRDVRTRP-EVGEMLRNKLVRL 386
            + ++ P+L++L          +            +L P  D R +  + G+  R  L+RL
Sbjct:  122 QNTLPPILAILLSLLSFFNIKQNL--------SMLLFPTNDDRKQSLQKGKSFRCLLLRL 173

Query:  387 MT-HLDTDVKRVAAEFLFVLCSESVPRFIKYTGYGNAAGLLAARGLMAGGRPEGQYS--- 442
            +T  +   +    A  L  LC     +  + G G A G+      M    P  + +
Sbjct:  174 LTIPIVEPIGTYYASLLNELCDGDSQQIARIFGAGYAMGISQHSETMPFPSPLSKAASPV 233

Query:  443 -EDEDTDTDEYKEAKASINPVTGRV--EEKPPNPMEGMTEEQKEHEAMKLVTMFDKLSRN 499
             +      +  +E   +I+P+TG +         +E +++E+KE EA +L  +F +L +N
Sbjct:  234 FQKNSRGQENTEENNLAIDPITGSMCTNRNKSQRLE-LSQEEKEREAERLFYLFQRLEKN 292
```

```
Query:  500 RVIQ 503
                IQ
Sbjct:  293 STIQ 296
```

Pedant information for DKFZphutel_22o2, frame 2

Report for DKFZphutel_22o2.2

```
[LENGTH]      537
[MW]          60372.53
[pI]          5.20
[BLOCKS]      BL00415L Synapsins proteins
[PROSITE]     MYRISTYL          4
[PROSITE]     CK2_PHOSPHO_SITE         13
[PROSITE]     PKC_PHOSPHO_SITE         10
[PROSITE]     ASN_GLYCOSYLATION        1
[KW]          All_Alpha
[KW]          LOW_COMPLEXITY      9.50 %
```

```
SEQ   MEPRAVAEAVETGEEDVIMEALRSYNQEHSQSFTFDDAQQEDRKRLAELLVSVLEQGLPP
SEG   ............................................................
PRD   ccchhhhhhhhhcccchhhhhhhhhhcccccccceeecccchhhhhhhhhhhhhhhcccccc

SEQ   SHRVIWLQSVRILSRDRNCLDPFTSRQSLQALACYADISVSEGSVPESADMDVVLESLKC
SEG   ............................................................
PRD   cceeeeecccccccccccccccccchhhhhhhhhhhceeeecccccccccchhhhhhhhh

SEQ   LCNLVLSSPVAQMLAAEARLVVKLTERVGLYRERSFPHDVQFFDLRLLFLLTALRTDVRQ
SEG   .........................................xxxxxxxxxxxxxxx...
PRD   hhhhccccchhhhhhhhhhhhhhhhhhcccccccccccccccchhhhhhhhhhhhhhhhhh

SEQ   QLFQELKGVRLLTDTLELTLGVTPEGNPPPTLLPSQETERAMEILKVLFNITLDSIKGEV
SEG   ............................................................
PRD   hhhhhhchhhhhhhhhhhhcccccccccccccccchhhhhhhhhhhhhhhhhhcccccchhh

SEQ   DEEDAALYRHLGTLLRHCVMIATAGDRTEEFHGHAVNLLGNLPLKCLDVLLTLEPHGDST
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhcccccccccccceeeeeccccccccceeeeeeeccccccc

SEQ   EFMGVNMDVIRALLIFLEKRLHKTHRLKESVAPVLSVLTECARMHRPARKFLKAQGWPPP
SEG   ............................................................
PRD   eeeehhhhhhhhhhhhhhhhhhhhhhhcccccceeeehhhhhhhhchhhhhhhhccccccc

SEQ   QVLPPLRDVRTRPEVGEMLRNKLVRLMTHLDTDVKRVAAEFLFVLCSESVPRFIKYTGYG
SEG   .........................................................xxx
PRD   ccccccccccchhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhcccccceeeecccc

SEQ   NAAGLLAARGLMAGGRPEGQYSEDEDTDTDEYKEAKASINPVTGRVEEKPPNPMEGMTEE
SEG   xxxxxxxxxxxxxxx.......xxxxxxxxx.............................
PRD   chhhhhhhhhcccccccccccccccccchhhhhhhhhccccccceeecccccccccchhhh

SEQ   QKEHEAMKLVTMFDKLSRNRVIQPMGMSPRGHLTSLQDAMCETMEQQLSSDPDSDPD
SEG   ................................................xxxxxxxxx
PRD   hhhhhhhhhhhhhhhcccccccccccccccccchhhhhhhhhhhhhhhhhhcccccccc
```

Prosite for DKFZphutel_22o2.2

PS00001	230->234	ASN_GLYCOSYLATION	PDOC00001
PS00005	61->64	PKC_PHOSPHO_SITE	PDOC00005
PS00005	69->72	PKC_PHOSPHO_SITE	PDOC00005
PS00005	84->87	PKC_PHOSPHO_SITE	PDOC00005
PS00005	117->120	PKC_PHOSPHO_SITE	PDOC00005
PS00005	145->148	PKC_PHOSPHO_SITE	PDOC00005
PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00005	324->327	PKC_PHOSPHO_SITE	PDOC00005
PS00005	463->466	PKC_PHOSPHO_SITE	PDOC00005
PS00005	508->511	PKC_PHOSPHO_SITE	PDOC00005
PS00006	12->16	CK2_PHOSPHO_SITE	PDOC00006
PS00006	34->38	CK2_PHOSPHO_SITE	PDOC00006
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	99->103	CK2_PHOSPHO_SITE	PDOC00006
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	263->267	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006

PS00006	388->392	CK2_PHOSPHO_SITE	PDOC00006
PS00006	442->446	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	530->534	CK2_PHOSPHO_SITE	PDOC00006
PS00008	57->63	MYRISTYL	PDOC00008
PS00008	420->426	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	430->436	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphutel_22o2.2)

DKFZphute1_23e13
----

group: metabolism

DKFZphtes3_15j18 encodes a novel 148 amino acid protein with similarity to 27K heat shock proteins.

The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. Subtilases are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases. Thus the novel protein is a new member of this family.

The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnologic production processes.

heat shock protein HSP27

strong similarity to heat shock 27K proteins

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: /map="578.9 cR from top of Chr12 linkage group"

Insert length: 1854 bp
Poly A stretch at pos. 1831, polyadenylation signal at pos. 1810

```
   1 GGTTTATTAA GCTCCTGGCT CCGCTCTAGA CCTCAGCGGT TCTGGCTGCC
  51 AGCCTGGGCA GCCTGGGAAG CCTGGGAGGA CGGTGGCTTG CCGGTCTGTC
 101 GTGAGGCAGT GCGGACGGGG ACCCTCTGGG ATTCTGCTGG ATCTGCCCCG
 151 GGGGTTACCT TTGGGGGCTG GGACCCCAGT CGAGGGGACA CAACCGTCCC
 201 TGGCAGTGGT TGGTTCTGCT TCTCCCTGCA GAAAAGCAGC ATTTTCGGAA
 251 GCTGAAGAAT AAGCTAGCCC AGCCACACCA CCTTGTTGTG TGACCTTGGG
 301 CAGGTGGTTC TGTCTCTCTG AGCCTCTGTT TCTCTCTGAG CTGAGCAGCC
 351 ACCATGGCTG ACGGTCAGAT GCCCTTCTCC TGCCACTACC CAAGCCGCCT
 401 GCGCCGAGAC CCCTTCCGGG ACTCTCCCCT CTCCTCTCGC CTGCTGGATG
 451 ATGGCTTTGG CATGGACCCC TTCCCAGACG ACTTGACAGC CTCTTGGCCC
 501 GACTGGGCTC TGCCTCGTCT CTCCTCCGCC TGGCCAGGCA CCCTAAGGTC
 551 GGGCATGGTG CCCCGGGGCC CCACTGCCAC CGCCAGGTTT GGGGTGCCTG
 601 CCGAGGGCAG GACCCCCCCA CCCTTCCCTG GGGAGCCCTG GAAAGTGTGT
 651 GTGAATGTGC ACAGCTTCAA GCCAGAGGAG TTGATGGTGA AGACCAAAGA
 701 TGGATACGTG GAGGTGTCTG GCAAACATGA AGAGAAACAG CAAGAAGGTG
 751 GCATTGTTTC TAAGAACTTC ACAAAGAAAA TCCAGCTTCC TGCAGAGGTG
 801 GATCCTGTGA CAGTATTTGC CTCACTTTCC CCAGAGGGTC TGCTGATCAT
 851 CGAAGCTCCC CAGGTCCCTC CTTACTCAAC ATTTGGAGAG AGCAGTTTCA
 901 ACAACGAGCT TCCCCAGGAC AGCCAGGAAG TCACCTGTAC CTGAGATGCC
 951 AGTACTGGCC CATCCTTGTT TTGTCCCCAA CCCTAGGGCT TCTCTGATTC
1001 CAGGATACAT TACTTTAGCT GAACTCAGAT TTAGTGCAAG TAAAATGTTA
1051 GAGGGTGCGG GGGTGAGGAC TGACCACAGA TTCCCTGGAT AGTGTAGTGG
1101 TAGATTTCTC CACAGGATAG CGCAATTGGC AAATCATGCT TGGTTGTGTT
1151 AGGCCAAAAT ACTAGTTTTG CTTTCTTTAC CTTTTCTATC TTGATGAAAA
1201 TGTTGCACAT TCTATAGTTG CAAAACACAT AAAAGGGGAC TTAACATTTC
1251 ACGTTGTATC TTACTTGCAG TGAATGCAAG GGTTACTTTT CTCTGGGGAC
1301 CTCCCCCATC ACCCAGGTTC CTACTCTGGG CTCCCGATTC CCATGGCTCC
1351 CAAACCATGC CGCATGGTTT GGTTAATGAA ACCCAGTAGC TAACCCCACT
1401 GTGCTTCCAC ATGCCTGGCC TAAAATGGGT GATATACAGG TCTTATATCC
1451 CCATATGGAA TTTATCCATC AACCACATAA AAACAAACAG TGCCTTCTGC
1501 CCTCTGCCCA GATGTGTCCA GCACGTTCTC AAAGTTTCCA CATTAGCACT
1551 CCCTAAGGAC GCTGGGAGCC TGTCAGTTTA TGATCTGACC TAGGTCCCCC
1601 CTTTCTTCTG TCCCCTGTGT TTAAGTCGGG ATTTTTACAG AGGGAGCTGT
1651 CTCCAGACAG CTCCATCAGG AACCAAGCAA AGGCCAGATA GCCTGACAGA
1701 TAGGCTAGTG GTATTGTGTA TATGGGCGGG ACGTGTGTGT CATTATTATT
1751 TGAGTTATGC TGTTGTTTAG GGGTAAATAA CAGTAAATAA TTAATAATAA
1801 TAATAATAAT AATAAAGGAG CTGACGTTCT TAAAAAAGAA AAAAAAAAAA
1851 AAAA
```

BLAST Results

Entry HS286348 from database EMBL:
human STS TIGR-A002J47.
Score = 510, P = 1.2e-16, identities = 102/102

Medline entries

95394379:
Cloning and sequencing of a cDNA encoding the canine HSP27 protein.

94110260:
Physiological and pathological changes in levels of the two small stress proteins, HSP27 and alpha B crystallin, in rat hindlimb muscles

Peptide information for frame 3

ORF from 354 bp to 941 bp; peptide length: 196
Category: strong similarity to known protein
Prosite motifs: SUBTILASE_ASP (28-39)

```
  1 MADGQMPFSC HYPSRLRRDP FRDSPLSSRL LDDGFGMDPF PDDLTASWPD
 51 WALPRLSSAW PGTLRSGMVP RGPTATARFG VPAEGRTPPP FPGEPWKVCV
101 NVHSFKPEEL MVKTKDGYVE VSGKHEEKQQ EGGIVSKNFT KKIQLPAEVD
151 PVTVFASLSP EGLLIIEAPQ VPPYSTFGES SFNNELPQDS QEVTCT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_23e13, frame 3

PIR:JC4244 heat-shock 27K protein - dog, N = 1, Score = 304, P = 4.3e-27

PIR:JN0924 heat shock 27 protein - rat, N = 1, Score = 301, P = 8.9e-27

TREMBL:MM03561_1 product: "heat shock protein HSP27"; Mus musculus heat shock protein HSP27 internal deletion variant b mRNA, complete cds., N = 1, Score = 301, P = 8.9e-27

>PIR:JC4244 heat-shock 27K protein - dog
Length = 209

HSPs:

Score = 304 (45.6 bits), Expect = 4.3e-27, P = 4.3e-27
Identities = 80/182 (43%), Positives = 102/182 (56%)

```
Query:    1 MADGQMPFSC-HYPSRLRRDPFRD-SPLSSRLLDDGFGMDPFPDDLTASWPDWALPRLSS 58
            M + ++PFS    PS     DPFRD  P  SRL D  FG+   P++     W  W     S
Sbjct:    1 MTERRVPFSLLRSPSW---DPFRDWYPAHSRLFDQAFGLPRLPEE----WAQWFG---HS 50

Query:   59 AWPGTLRSGMVP---RGPTATARFGVPAEGR--TPPPFPG--------EPWKVCVNVHSF 105
             WFG +R   +P    GP A A    PA  R  +     G         + W+V ++V+ F
Sbjct:   51 GWPGYVRP--IPPAVEGPAAAAAAAAPAYSRALSRQLSSGVSEIRQTADRWRVSLDVNHF 108

Query:  106 KPEELMVKTKDGYVEVSGKHEEKQQEGGIVSKNFTKKIQLPAEVDPVTVFASLSPEGLLI 165
             PEEL VKTKDG VE++GKHEE+Q E G +S+  T K  LP  VDP  V +SLSPEG L
Sbjct:  109 APEELTVKTKDGVVEITGKHEERQDEHGYISRRLTPKYTLPPGVDPTLVSSSLSPEGTLT 168

Query:  166 IEAPQVPPYSTFGE 179
            +EAP   P +   E
Sbjct:  169 VEAPMPKPATQSAE 182
```

Pedant information for DKFZphutel_23e13, frame 3

Report for DKFZphutel_23e13.3

[LENGTH] 196
[MW] 21604.37

```
[pI]            5.00
[HOMOL]         PIR:JC4244 heat-shock 27K protein - dog 3e-22
[BLOCKS]        BL01031C
[PIRKW]         blocked amino end 1e-13
[PIRKW]         acetylated amino end 4e-13
[PIRKW]         phosphoprotein 7e-21
[PIRKW]         glycoprotein 2e-11
[PIRKW]         heat shock 7e-21
[PIRKW]         molecular chaperone 4e-13
[PIRKW]         alternative splicing 1e-19
[PIRKW]         eye lens 6e-14
[PIRKW]         stress-induced protein 7e-21
[SUPFAM]        alpha-crystallin 7e-21
[PROSITE]       SUBTILASE_ASP  1
[PROSITE]       MYRISTYL       2
[PROSITE]       CK2_PHOSPHO_SITE      2
[PROSITE]       PKC_PHOSPHO_SITE      6
[PROSITE]       ASN_GLYCOSYLATION     1
[PFAM]          Heat shock hsp20 proteins
[KW]            All_Beta
[KW]            LOW_COMPLEXITY     7.14 %

SEQ    MADGQMPFSCHYPSRLRRDPFRDSPLSSRLLDDGFGMDPFPDDLTASWPDWALPRLSSAW
SEG    ............................xxxxxxxxxxxxxx................
PRD    cccccccccccccccccccccccccccchhhhhcccccccccccccccccccccccccccc

SEQ    PGTLRSGMVPRGPTATARFGVPAEGRTPPPFPGEPWKVCVNVHSFKPEELMVKTKDGYVE
SEG    ............................................................
PRD    ccccccccccccchhhhhhhhcccccchhhhhheeeeeecccccceeeeeeccccceee

SEQ    VSGKHEEKQQEGGIVSKNFTKKIQLPAEVDPVTVFASLSPEGLLIIEAPQVPPYSTFGES
SEG    ............................................................
PRD    ecchhhhhccccceeeecccccccccccccccceeeeccccccceeeeecccccccccccc

SEQ    SFNNELPQDSQEVTCT
SEG    ................
PRD    cccccccccceeeccc
```

Prosite for DKFZphute1_23e13.3

PS00001	138->142	ASN_GLYCOSYLATION	PDOC00001
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	140->143	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	62->68	MYRISTYL	PDOC00008
PS00008	132->138	MYRISTYL	PDOC00008
PS00136	28->39	SUBTILASE_ASP	PDOC00125

Pfam for DKFZphute1_23e13.3

```
HMM_NAME     Heat shock hsp20 proteins

HMM              *AMMrpPWDWRE.....DpDHFeVrMDMPGFKPEEIKVkVEDNNVLvIeG
                  A    P++ R        + ++V++++  FKPEE+ VK+ D+ +++++G
Query         77  ARFGVPAEGR-TPPPFPGEPWKVCVNVHSFKPEELMVKTKDG-YVEVSG    123

HMM              EHEREEEREDDkWWWHERIYRHFMRRFrLPENVDpDqIkAsMSdNGVLTI
                 +HE    E++      + + ++  F  +++LP +VDP  + AS+S++G+L I
Query        124 KHE---EKQQ----EGGIVSKNFTKKIQLPAEVDPVTVFASLSPEGLLII    166

HMM              TVPKpEP*
                 ++P ++P
Query        167 EAPQVPP     173
```

DKFZphutel_23g11

group: uterus derived

DKFZphutel_23g11 encodes a novel 256 amino acid protein with similarity to S.pombe SPAC31G5.12c and S. cerevisiae Maf1p.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to SPAC31G5.12c and Maf1p

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1674 bp
Poly A stretch at pos. 1664, polyadenylation signal at pos. 1644

```
   1 GGGGGAGGCG GAGGTCGCTC GCTCGCTCGC TCGGCTCGCT GACTCGCCGG
  51 AGCGCTCTGT GGCGGTCGGC GGCAGGTCGG TCGCGAGAGC GGGCTCTGTG
 101 GAAGGGGGCG AGGCTATGTC GCGGTGGCAG CCCGGATGGG CCGGCAGGGC
 151 CGGGAGTAAC GGGACGTCGC CGCGGAGCTT CTTCCCCCGG ATACAGTGCG
 201 GCCCGAGCGG AGGCCGCGGC GCCGCCCTCC GATCTTGAAG AGCCCGCGCT
 251 GCGCGGAGCC CGCCCCCGCC TGCGCACCGG CACCGACGCG GAGCGACCAG
 301 CCCAGCCAGA CCCGGCCCGG CGCGGCCTGA TCTAACCCAG CCAGGCAGGC
 351 AATACTAGCC CCTCTGGAGC ACGGAGCTCC TTCCCCAAAG ACATGAAGCT
 401 ATTGGAGAAC TCGAGCTTTG AAGCCATCAA CTCACAGCTG ACTGTGGAGA
 451 CCGGAGATGC CCACATCATT GGCAGGATTG AGAGCTACTC ATGTAAGATG
 501 GCAGGAGACG ACAAACACAT GTTCAAGCAG TTCTGCCAGG AGGGCCAGCC
 551 CCACGTGCTG GAGGCACTTT CTCCACCCCA GACTTCAGGA CTGAGCCCCA
 601 GCAGACTCAG CAAAAGCCAA GGCGGTGAGG AGGAGGGCCC CCTCAGTGAC
 651 AAGTGCAGCC GCAAGACCCT CTTCTACCTG ATTGCCACGC TCAATGAGTC
 701 CTTCAGGCCT GACTATGACT TCAGCACAGC CCGCAGCCAT GAGTTCAGCC
 751 GGGAGCCCAG CCTTAGCTGG GTGGTGAATG CAGTCAACTG CAGTCTGTTC
 801 TCAGCTGTGC GGGAGGACTT CAAGGATCTG AAACCACAGC TGTGGAACGC
 851 GGTGGACGAG GAGATCTGCC TGGCTGAATG TGACATCTAC AGCTATAACC
 901 CAGACTTGGA CTCAGATCCC TTCGGGGAGG ATGGTAGCCT CTGGTCCTTC
 951 AACTACTTCT TCTACAACAA GCGGCTCAAG CGAATCGTCT TCTTTAGCTG
1001 CCGTTCCATC AGTGGCTCCA CCTACACACC CTCAGAGGCA GGCAACGAGC
1051 TGGACATGGA GCTGGGGGAG GAGGAGGTGG AGGAAGAAAG CAGAAGCAGG
1101 GGCAGTGGGG CCGAGGAGAC CAGCACCATG GAGGAGGACA GGGTCCCAGT
1151 GATCTGTATT TGATGAGGAG GAGCCGAGGC CCCAGCTTCA TCCAGCTTCA
1201 ACCAATGCCT GGACCTGTCC ACCTGAGAGG CCCCTGGGGC CTCCCCAGCT
1251 GCTGGCCAGA CCCTGGCGCT GCCACAGTCC TGGCACTGCC CAAGGCCATA
1301 CCTGCCTAGC CCTTTGGCTC CATCCTGTGG ATGCCCACTC ACCCCTCAGA
1351 CTCCTGCTGC CCATGCTGTG GCCGGACTTG TCAGCAGGGG GCCTGGTGGG
1401 AGGAGCGACT GCCCTGCCCA AATGAACTGC CACAGCAGGG ACAGCTGGAC
1451 CGCAGAGTTT ATTTTTGTAT TTCTACTGGG CCTGCACACT CCAGCCCAAA
1501 GGGTCTGTGG CCGGAGGCCC CACGAGCAGG CCCCAGCAGT CACCGGCTCT
1551 GGTCTTGGGC CGGCCCCGGT GCCCACCTGT ACCCCCACCT CGCCCATTTG
1601 GCCGCGTGCA CTGAGTGTCA CTTTGCTGCA GCTCGTTTCT TTCCAATAAA
1651 AGTTTCTGTG ACTTAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 393 bp to 1160 bp; peptide length: 256
Category: similarity to known protein

```
  1 MKLLENSSFE AINSQLTVET GDAHIIGRIE SYSCKMAGDD KHMFKQFCQE
 51 GQPHVLEALS PPQTSGLSPS RLSKSQGGEE EGPLSDKCSR KTLFYLIATL
101 NESFRPDYDF STARSHEFSR EPSLSWVVNA VNCSLFSAVR EDFKDLKPQL
151 WNAVDEEICL AECDIYSYNP DLDSDPFGED GSLWSFNYFF YNKRLKRIVF
201 FSCRSISGST YTPSEAGNEL DMELGEEEVE EESRSRGSGA EETSTMEEDR
251 VPVICI
```

BLASTP hits

Entry SPAC31G5_12 from database TREMBL:
gene: "SPAC31G5.12c"; product: "hypothetical protein"; S.pombe chromosome I cosmid c31G5.
Score = 272, P = 9.3e-24, identities = 51/127, positives = 80/127

Entry SPD656_1 from database TREMBL:
product: "ORF N150"; Yeast DNA for bfr2+ protein/pad1+ protein/sks1+ protein, ORF N313, ORF N150, complete cds, and for ORF N118, partial cds.
Score = 263, P = 8.4e-23, identities = 50/127, positives = 79/127

Entry S50986 from database PIR:
MAF1 protein - yeast (Saccharomyces cerevisiae) >SWISSPROT:MAF1_YEAST MAF1 PROTEIN. >TREMBL:SC19492_1 gene: "MAF1"; product: "Maf1p"; Saccharomyces cerevisiae Maf1p (MAF1) gene, complete cds. >TREMBL:SC8119_11 gene: "MAF1p"; product: "Maf1p"; S.cerevisiae chromosome IV cosmid 8119.
Score = 180, P = 2.3e-17, identities = 43/133, positives = 75/133

Entry AF098499_2 from database TREMBL:
gene: "C43H8.2"; Caenorhabditis elegans cosmid C43H8.
Score = 263, P = 9.2e-23, identities = 78/252, positives = 118/252

Alert BLASTP hits for DKFZphutel_23g11, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphutel_23g11, frame 3

Report for DKFZphutel_23g11.3

```
[LENGTH]        256
[MW]            28869.95
[pI]            4.51
[HOMOL]         TREMBL:SPAC31G5_12 gene: "SPAC31G5.12c"; product: "hypothetical protein";
S.pombe chromosome I cosmid c31G5. 4e-23
[FUNCAT]        06.04 protein targeting, sorting and translocation   [S. cerevisiae, YDR005c]
6e-13
[PROSITE]       MYRISTYL        3
[PROSITE]       CK2_PHOSPHO_SITE        5
[PROSITE]       PKC_PHOSPHO_SITE        6
[PROSITE]       ASN_GLYCOSYLATION       3
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY     7.81 %

SEQ   MKLLENSSFEAINSQLTVETGDAHIIGRIESYSCKMAGDDKHMFKQFCQEGQPHVLEALS
SEG   ............................................................
PRD   cccccchhhhhhhhhhhhccccceeeeeccccchhhhhccchhhhhhhhhccccceeeeccc

SEQ   PPQTSGLSPSRLSKSQGGEEEGPLSDKCSRKTLFYLIATLNESFRPDYDFSTARSHEFSR
SEG   ............................................................
PRD   cccccccccccccccccccccccccccchhhhhhhhhhhccccccccccccccccccc

SEQ   EPSLSWVVNAVNCSLFSAVREDFKDLKPQLWNAVDEEICLAECDIYSYNPDLDSDPFGED
SEG   ............................................................
PRD   cccccchhhhhhhhhhhhhchhhhhhhhhhhhhhhhcccccceeecccccccccccc

SEQ   GSLWSFNYFFYNKRLKRIVFFSCRSISGSTYTPSEAGNELDMELGEEEVEEESRSRGSGA
SEG   ..........................................xxxxxxxxxxxxxxxxxx
PRD   ccceeeceeechhhhhhhhhhhcccccccccccccccchhhhhhhhhhhhhcccccc

SEQ   EETSTMEEDRVPVICI
SEG   xx..............
PRD   cccccccccceeeccc
```

Prosite for DKFZphutel_23g11.3

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00001	132->136	ASN_GLYCOSYLATION	PDOC00001
PS00005	33->36	PKC_PHOSPHO_SITE	PDOC00005
PS00005	85->88	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	202->205	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	99->103	CK2_PHOSPHO_SITE	PDOC00006
PS00006	212->216	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	181->187	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphutel_23g11.3)

DKFZphutel_24c19

group: transmembrane protein

DKFZphutel_24c19 encodes a novel 195 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

unknown
membrane regions: 1
Summary DKFZphutel_24c19 encodes a novel 195 amino acid protein, with no similarity to known proteins.

unknown

complete cDNA, complete cds, EST hits
TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 769 bp
Poly A stretch at pos. 746, polyadenylation signal at pos. 735

```
  1 ACGAGTCAGC CAAAGATGGC TGCGCCCAGG TAATTTGAGC AAAGGCCACA
 51 GTGAACTCCG GCGTGGCTGA GGAAGACCGG AGGAGGCACC CACAGGCTGC
101 TGGGAGGAGA GCATAAGGCT CAAAATGGAA AATCATAAAT CCAATAATAA
151 GGAAAACATA ACAATTGTTG ATATATCCAG AAAAATTAAC CAGCTTCCAG
201 AAGCAGAAAG GAATCTACTT GAAAATGGAT CGGTTTATGT TGGATTAAAT
251 GCTGCTCTTT GTGGCCTCAT AGCAAACAGT CTTTTTCGAC GCATCTTGAA
301 TGTGACAAAG GCTCGCATAG CTGCTGGCTT ACCAATGGCA GGGATACCTT
351 TTCTTACAAC AGACTTAACT TACAGATGTT TTGTAAGTTT TCCTTTGAAT
401 ACAGGTGATT TGGATTGTGA AACCTGTACC ATAACACGGA GTGGACTGAC
451 TGGTCTTGTT ATTGGTGGTC TATACCCTGT TTTCTTGGCT ATACCTGTAA
501 ATGGTGGTCT AGCAGCCAGG TATCAATCAG CTCTGTTACC ACACAAAGGG
551 AACATCTTAA GTTACTGGAT TAGAACTTCT AAGCCTGTCT TTAGAAAGAT
601 GTTATTTCCT ATTTTGCTCC AGACTATGTT TTCAGCATAC CTTGGGTCTG
651 AACAATATAA ACTACTTATA AAGGCCCTTC AGTTATCTGA ACCTGGCAAA
701 GAAATTCACT GATTTTAAAC AAATATGTAA ACAAAAATAA AATGGTAAAA
751 ACAAAAAAAA AAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 125 bp to 709 bp; peptide length: 195
Category: putative protein

```
  1 MENHKSNNKE NITIVDISRK INQLPEAERN LLENGSVYVG LNAALCGLIA
 51 NSLFRRILNV TKARIAAGLP MAGIPFLTTD LTYRCFVSFP LNTGDLDCET
101 CTITRSGLTG LVIGGLYPVF LAIPVNGGLA ARYQSALLPH KGNILSYWIR
151 TSKPVFRKML FPILLQTMFS AYLGSEQYKL LIKALQLSEP GKEIH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_24c19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphutel_24c19, frame 2

Report for DKFZphutel_24c19.2

[LENGTH]	195	
[MW]	21527.45	
[pI]	9.36	
[PROSITE]	MYRISTYL	6
[PROSITE]	CK2_PHOSPHO_SITE	1
[PROSITE]	PKC_PHOSPHO_SITE	3
[PROSITE]	ASN_GLYCOSYLATION	3
[KW]	TRANSMEMBRANE	1

```
SEQ  MENHKSNNKENITIVDISRKINQLPEAERNLLENGSVYVGLNAALCGLIANSLFRRILNV
PRD  cccccccccceeeeeehhhhhhcchhhhhhhcccccceeeecchhhhhhhhhhhhhhhh
MEM  ............................................................

SEQ  TKARIAAGLPMAGIPFLTTDLTYRCFVSFPLNTGDLDCETCTITRSGLTGLVIGGLYPVF
PRD  hhhhhhhcccccccceeeeecccccccccccccccccccccccccccccceeeecccceee
MEM  .............................................MMMMMMMMMMMMMMM

SEQ  LAIPVNGGLAARYQSALLPHKGNILSYWIRTSKPVFRKMLFPILLQTMFSAYLGSEQYKL
PRD  eeeccccccchhhhhhccccccceeeeeeeccccchhhhhchhhhhhhhhhhhhcchhhhh
MEM  MMM.........................................................

SEQ  LIKALQLSEPGKEIH
PRD  hhhhhhhcccccccc
MEM  ...............
```

Prosite for DKFZphutel_24c19.2

PS00001	11->15	ASN_GLYCOSYLATION	PDOC00001
PS00001	34->38	ASN_GLYCOSYLATION	PDOC00001
PS00001	59->63	ASN_GLYCOSYLATION	PDOC00001
PS00005	18->21	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	151->154	PKC_PHOSPHO_SITE	PDOC00005
PS00006	13->17	CK2_PHOSPHO_SITE	PDOC00006
PS00008	40->46	MYRISTYL	PDOC00008
PS00008	47->53	MYRISTYL	PDOC00008
PS00008	68->74	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008
PS00008	142->148	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphutel_24c19.2)

DKFZphutel_24e11
----

group: intracellular transport and trafficking

DKFZphutel_24e11 encodes a novel 226 amino acid protein, with similarity to human/mouse golgi 4-transmembrane spanning transporter MTP. MTP may function in the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartment. Thus, the novel protein also seems to be involved in nucleotide sugar transport.

The new protein can find application in modulating the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartments.

similarity to 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP

complete cDNA, complete cds, EST hits
potential start at 184,
TRANSMEMBRANE 4
function in the transport of nucleosides and/or nucleoside derivatives
between the cytosol and
the lumen of an intracellular membrane-bound compartment?

Sequenced by Qiagen

Locus: /map="8"

Insert length: 2005 bp
Poly A stretch at pos. 1988, polyadenylation signal at pos. 1963

```
   1 ACGCGTCCGG CAGAAGCTCG GAGCTCTCGG GGTATCGAGG AGGCAGGCCC
  51 GCGGGCGCAC GGGCGAGCGG GCCGGGAGCC GGAGCGGCGG AGGAGCCGGC
 101 AGCAGCGGCG CGGCGGGCTC CAGGCGAGGC GGTCGACGCT CCTGAAAACT
 151 TGCGCGCGCG CTCGCGCCAC TGCGCCCGGA GCGATGAAGA TGGTCGCGCC
 201 CTGGACGCGG TTCTACTCCA ACAGCTGCTG CTTGTGCTGC CATGTCCGCA
 251 CCGGCACCAT CCTGCTCGGC GTCTGGTATC TGATCATCAA TGCTGTGGTA
 301 CTGTTGATTT TATTGAGTGC CCTGGCTGAT CCGGATCAGT ATAACTTTTC
 351 AAGTTCTGAA CTGGGAGGTG ACTTTGAGTT CATGGATGAT GCCAACATGT
 401 GCATTGCCAT TGCGATTTCT CTTCTCATGA TCCTGATATG TGCTATGGCT
 451 ACTTACGGAG CGTACAAGCA ACGCGCAGCC TGGATCATCC CATTCTTCTG
 501 TTACCAGATC TTTGACTTTG CCCTGAACAT GTTGGTTGCA ATCACTGTGC
 551 TTATTTATCC AAACTCCATT CAGGAATACA TACGGCAACT GCCTCCTAAT
 601 TTTCCCTACA GAGATGATGT CATGTCAGTG AATCCTACCT GTTTGGTCCT
 651 TATTATTCTT CTGTTTATTA GCATTATCTT GACTTTTAAG GGTTACTTGA
 701 TTAGCTGTGT TTGGAACTGC TACCGATACA TCAATGGTAG GAACTCCTCT
 751 GATGTCCTGG TTTATGTTAC CAGCAATGAC ACTACGGTGC TGCTACCCCC
 801 GTATGATGAT GCCACTGTGA ATGGTGCTGC CAAGGAGCCA CCGCCACCTT
 851 ACGTGTCTGC CTAAGCCTTC AAGTGGGCGG AGCTGAGGGC AGCAGCTTGA
 901 CTTTGCAGAC ATCTGAGCAA TAGTTCTGTT ATTTCACTTT TGCCATGAGC
 951 CTCTCTGAGC TTGTTTGTTG CTGAAATGCT ACTTTTTAAA ATTTAGATGT
1001 TAGATTGAAA ACTGTAGTTT TCAACATATG CTTTGCTAGA ACACTGTGAT
1051 AGATTAACTG TAGAATTCTT CCTGTACGAT TGGGGATATA ACGGGCTTCA
1101 CTAACCTTCC CTAGGCATTG AAACTTCCCC CAAATCTGAT GGACCTAGAA
1151 GTCTGCTTTT GTACCTGCTG GGCCCCAAAG TTGGGCATTT TTCTCTCTGT
1201 TCCCTCTCTT TTGAAAATGT AAAATAAAAC CAAAAATAGA CAACTTTTTC
1251 TTCAGCCATT CCAGCATAGA GAACAAAACC TTATGGAAAC AGGAATGTCA
1301 ATTGTGTAAT CATTGTTCTA ATTAGGTAAA TAGAAGTCCT TATGTATGTG
1351 TTACAAGAAT TTCCCCCACA ACATCCTTTA TGACTGAAGT TCAATGACAG
1401 TTTGTGTTTG GTGGTAAAGG ATTTTCTCCA TGGCCTGAAT TAAGACCATT
1451 AGAAAGCACC AGGCCGTGGG AGCAGTGACC ATCTACTGAC TGTTCTTGTG
1501 GATCTTGTGT CCAGGGACAT GGGGTGACAT GCCTCGTATG TGTTAGAGGG
1551 TGGAATGGAT GTGTTTGGCG CTGCATGGGA TCTGGTGCCC CTCTTCTCCT
1601 GGATTCACAT CCCCACCCAG GGCCCGCTTT TACTAAGTGT TCTGCCCTAG
1651 ATTGGTTCAA GGAGGTCATC CAACTGACTT TATCAAGTGG AATTGGGATA
1701 TATTTGATAT ACTTCTGCCT AACAACATGG AAAAGGGTTT TCTTTTCCCT
1751 GCAAGCTACA TCCTACTGCT TTGAACTTCC AAGTATGTCT AGTCACCTTT
1801 TAAAATGTAA ACATTTTCAG AAAAATGAGG ATTGCCTTCC TTGTATGCGC
1851 TTTTTACCTT GACTACCTGA ATTGCAAGGG ATTTTTATAT ATTCATATGT
1901 TACAAAGTCA GCAACTCTCC TGTTGGTTCA TTATTGAATG TGCTGTAAAT
1951 TAAGTCGTTT GCAATTAAAA CAAGGTTTGC CCACATCCAA AAAAAAAAAA
2001 AAAAA
```

BLAST Results

Entry HS012351 from database EMBL:

human STS SHGC-31823.
Score = 1629, P = 3.1e-67, identities = 343/354

Medline entries

96199248:
Identification of a novel membrane transporter associated with intracellular membranes by phenotypic complementation in the yeast Saccharomyces cerevisiae.

Peptide information for frame 1

ORF from 184 bp to 861 bp; peptide length: 226
Category: strong similarity to known protein

```
  1 MKMVAPWTRF YSNSCCLCCH VRTGTILLGV WYLIINAVVL LILLSALADP
 51 DQYNFSSSEL GGDFEFMDDA NMCIAIAISL LMILICAMAT YGAYKQRAAW
101 IIPFFCYQIF DFALNMLVAI TVLIYPNSIQ EYIRQLPPNF PYRDDVMSVN
151 PTCLVLIILL FISIILTFKG YLISCVWNCY RYINGRNSSD VLVYVTSNDT
201 TVLLPPYDDA TVNGAAKEPP PPYVSA
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_24e11, frame 1

SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP (KIAA0108)., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:MTRP_MOUSE GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP., N = 1, Score = 539, P = 5.3e-52

TREMBL:HS304981_1 product: "E3 protein"; Human retinoic acid-inducible E3 protein mRNA, complete cds., N = 1, Score = 127, P = 3.4e-06

```
>SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP
     (KIAA0108).
           Length = 233

  HSPs:

 Score = 551 (82.7 bits), Expect = 2.9e-53, P = 2.9e-53
 Identities = 102/221 (46%), Positives = 148/221 (66%)

Query:      9 RFYSNSCCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQY---NFSSSELGGDF- 64
              RFYS  CC CCHVRTGTI+LG WY+++N ++ ++L   +  P+     N     +G  +
Sbjct:     13 RFYSTRCCGCCHVRTGTIILGTWYMVVNLLMAILLTVEVTHPNSMPAVNIQYEVIGNYYS 72

Query:     65 -EFMDDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAITVL 123
               E M D N C+  A+S+LM +I +M  YGA    +  W+IPFFCY++FDF L+ LVAI+ L
Sbjct:     73 SERMAD-NACVLFAVSVLMFIISSMLVYGAISYQVGWLIPFFCYRLFDFVLSCLVAISSL 131

Query:    124 IYPNSIQEYIRQLPPNFPYRDDVMSVNPTCLVLIILLFISIILTFKGYLISCVWNCYRYI 183
              Y   I+EY+ QLP +FPY+DD+++++ +CL+ I+L+F ++ + FK YLI+CVWNCY+YI
Sbjct:    132 TYLPRIKEYLDQLP-DFPYKDDLLALDSSCLLFIVLVFFALFIIFKAYLINCVWNCYKYI 190

Query:    184 NGRNSSDVLVYVTSN-DTTVLLPPYDDATVNGAAKEPPPPYVSA 226
              N RN  ++ VY          +LP Y+ A V    KEPPPPY+ A
Sbjct:    191 NNRNVPEIAVYPAFEAPPQYVLPTYEMA-VKMPEKEPPPPYLPA 233
```

Pedant information for DKFZphutel_24e11, frame 1

Report for DKFZphutel_24e11.1

[LENGTH]	226
[MW]	25419.11

```
[pI]            4.65
[HOMOL]         SWISSPROT:MTRP_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP (KIAA0108).
5e-40
[PROSITE]       CK2_PHOSPHO_SITE      3
[PROSITE]       TYR_PHOSPHO_SITE      1
[PROSITE]       PKC_PHOSPHO_SITE      1
[PROSITE]       ASN_GLYCOSYLATION     3
[KW]            SIGNAL_PEPTIDE 49
[KW]            TRANSMEMBRANE  2
[KW]            LOW_COMPLEXITY    20.80 %

SEQ     MKMVAPWTRFYSNSCCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQYNFSSSEL
SEG     ................................xxxxxxxxxxxxxxxx............
PRD     ccceeeeeeeccccceeeeeeeeecccceeeecceeehhhhhhhhhhhhhhcccccceeecccc
MEM     ............................................................

SEQ     GGDFEFMDDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAI
SEG     ...........xxxxxxxxxxxxxxxxxx...............................
PRD     cccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhh
MEM     .........MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ     TVLIYPNSIQEYIRQLPPNFPYRDDVMSVNPTCLVLIIILLFISIILTFKGYLISCVWNCY
SEG     ................................xxxxxxxxxxxxx..............
PRD     hhhccchhhhhhhhccccccccccceeeecccccceeehhhhhhhhhhhhhheeeeeeeee
MEM     MMMMMM....................MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM....

SEQ     RYINGRNSSDVLVYVTSNDTTVLLPPYDDATVNGAAKEPPPPYVSA
SEG     ..............................................
PRD     eeccccccccceeeeeecccccccccccccccccccccccccccccc
MEM     ..............................................
```

Prosite for DKFZphutel_24e11.1

PS00001	54->58	ASN_GLYCOSYLATION	PDOC00001
PS00001	187->191	ASN_GLYCOSYLATION	PDOC00001
PS00001	198->202	ASN_GLYCOSYLATION	PDOC00001
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00006	56->60	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00007	186->195	TYR_PHOSPHO_SITE	PDOC00007

(No Pfam data available for DKFZphutel_24e11.1)

DKFZphutel_24j6
----

group: cell structure and motility

DKFZphutesl_24j6 encodes a novel 571 amino acid protein with strong similarity to rat cell adhesion regulator (CAR1).

The novel protein is very similar to Carl and thus seems to be involved in regulation cell-cell adhesion. It contains a RGD cell attachment site.

The new protein can find application in modulation of cell-cell-adhesion.

strong similarity to rat CAR1 A.thaliana T19C21.5

complete cDNA, complete cds, EST hits
potential frame shift at Bp 1241 according to CAR1
but frame shift might be in CAR1 sequence!
ESTs T73366 AA362984 confirm this sequence

Sequenced by Qiagen

Locus: /map="939.9 cR from top of Chr2 linkage group"

Insert length: 3333 bp
Poly A stretch at pos. 3316, no polyadenylation signal found

```
   1 ACGCGTCCGA GCTGGCTCAG GGCGTCCGCT AGGCTCGGAC GACCTGCTGA
  51 GCCTCCCAAA CCGCTTCCAT AAGGCTTTGC CTTTCCAACT TCAGCTACAG
 101 TGTTAGCTAA GTTTGGAAAG AAGGAAAAAA GAAAATCCCT GGGCCCCTTT
 151 TCTTTTGTTC TTTGCCAAAG TCGTCGTTGT AGTCTTTTTG CCCAAGGCTG
 201 TTGTGTTTTT AGAGGTGCTA TCTCCAGTTC CTTGCACTCC TGTTAACAAG
 251 CACCTCAGCG AGAGCAGCAG CAGCGATAGC AGCCGCAGAA GAGCCAGCGG
 301 GGTCGCCTAG TGTCATGACC AGGGCGGGAG ATCACAACCG CCAGAGAGGA
 351 TGCTGTGGAT CCTTGGCCGA CTACCTGACC TCTGCAAAAT TCCTTCTCTA
 401 CCTTGGTCAT TCTCTCTCTA CTTGGGGAGA TCGGATGTGG CACTTTGCGG
 451 TGTCTGTGTT TCTGGTAGAG CTCTATGGAA ACAGCCTCCT TTTGACAGCA
 501 GTCTACGGGC TGGTGGTGGC AGGGTCTGTT CTGGTCCTGG GAGCCATCAT
 551 CGGTGACTGG GTGGACAAGA ATGCTAGACT TAAAGTGGCC CAGACCTCGC
 601 TGGTGGTACA GAATGTTTCA GTCATCCTGT GTGGAATCAT CCTGATGATG
 651 GTTTTCTTAC ATAAACATGA GCTTCTGACC ATGTACCATG GATGGGTTCT
 701 CACTTCCTGC TATATCCTGA TCATCACTAT TGCAAATATT GCAAATTTGG
 751 CCAGTACTGC TACTGCAATC ACAATCCAAA GGGATTGGAT TGTTGTTGTT
 801 GCAGGAGAAG ACAGAAGCAA ACTAGCAAAT ATGAATGCCA CAATACGAAG
 851 GATTGACCAG TTAACCAACA TCTTAGCCCC CATGGCTGTT GGCCAGATTA
 901 TGACATTTGG CTCCCCAGTC ATCGGCTGTG GCTTTATTTC GGGATGGAAC
 951 TTGGTATCCA TGTGCGTGGA GTACGTCCTG CTCTGGAAGG TTTACCAGAA
1001 AACCCCAGCT CTAGCTGTGA AAGCTGGTCT TAAAGAAGAG GAAACTGAAT
1051 TGAAACAGCT GAATTTACAC AAAGATACTG AGCCAAAACC CCTGGAGGGA
1101 ACTCATCTAA TGGGTGTGAA AGACTCTAAC ATCCATGAGC TTGAACATGA
1151 GCAAGAGCCT ACTTGTGCCT CCCAGATGGC TGAGCCCTTC CGTACCTTCC
1201 GAGATGGATG GGTCTCCTAC TACAACCAGC CTGTGTTTCT GGCTGGCATG
1251 GGTCTTGCTT TCCTTTATAT GACTGTCCTG GGCTTTGACT GCATCACCAC
1301 AGGGTACGCC TACACTCAGG GACTGAGTGG TTCCATCCTC AGTATTTTGA
1351 TGGGAGCATC AGCTATAACT GGAATAATGG GAACTGTAGC TTTTACTTGG
1401 CTACGTCGAA AATGTGGTTT GGTTCGGACA GGTCTGATCT CAGGATTGGC
1451 ACAGCTTTCC TGTTTGATCT TGTGTGTGAT CTCTGTATTC ATGCCTGGAA
1501 GCCCCCTGGA CTTGTCCGTT TCTCCTTTTG AAGATATCCG ATCAAGGTTC
1551 ATTCAAGGAG AGTCAATTAC ACCTACCAAG ATACCTGAAA TTACAACTGA
1601 AATATACATG TCTAATGGGT CTAATTCTGC TAATATTGTC CCGGAGACAA
1651 GTCCTGAATC TGTGCCCATA ATCTCTGTCA GTCTGCTGTT TGCAGGCGTC
1701 ATTGCTGCTA GAATCGGTCT TTGGTCCTTT GATTTAACTG TGACACAGTT
1751 GCTGCAAGAA AATGTAATTG AATCTGAAAG AGGCATTATA AATGGTGTAC
1801 AGAACTCCAT GAACTATCTT CTTGATCTTC TGCATTTCAT CATGGTCATC
1851 CTGGCTCCAA ATCCTGAAGC TTTTGGCTTG CTCGTATTGA TTTCAGTCTC
1901 CTTTGTGGCA ATGGGCCACA TTATGTATTT CCGATTTGCC CAAAATACTC
1951 TGGGAAACAA GCTCTTTGCT TGCGGTCCTG ATGCAAAAGA AGTTAGGAAG
2001 GAAAATCAAG CAAATACATC TGTTGTTTGA GACAGTTTAA CTGTTGCTAT
2051 CCTGTTACTA GATTATATAG AGCACATGTG CTTATTTTGT ACTGCAGAAT
2101 TCCAATAAAT GGCTGGGTGT TTTGCTCTGT TTTTACCACA GCTGTGCCTT
2151 GAGAACTAAA AGCTGTTTAG GAAACCTAAG TCAGCAGAAA TTAACTGATT
2201 AATTTCCCTT ATGTTGAGGC ATGGAAAAAA AATTGGAAAA GAAAAACTCA
2251 GTTTAAATAC GGAGACTATA ATGATAACAC TGAATTCCCC TATTTCTCAT
2301 GAGTAGATAC AATCTTACGT AAAAGAGTGG TTAGTCACGT GAATTCAGTT
2351 ATCATTTGAC AGATTCTTAT CTGTACTAGA ATTCAGATAT GTCAGTTTTC
2401 TGCAAAACTC ACTCTTGTTC AAGACTAGCT AATTTATTTT TTTGCATCTT
2451 AGTTATTTTT AAAAACAAAT TCTTCAAGTA TGAAGACTAA ATTTGATAA
2501 CTAATATTAT CCTTATTGAT CCTATTGATC TTAAGGTATT TACATGTATG
```

```
2551 TGGAAAAACA AAACACTTAA CTAGAATTCT CTAATAAGGT TTATGGTTTA
2601 GCTTAAAGAG CACCTTTGTA TTTTTATTAT CAGATGGGGC AACATATTGT
2651 ATGAAGCATA TGTAGCACTT CACAGCATGG TTATCATGTA AGCTGCAGGT
2701 AGAAGCAAAG CTGTAAAGTA GATTTATCAC ACAATGACTG CATACAGACT
2751 TCAAATATGT CAATAGTTTG GTCATAGAAC CTAGAAGCCA AAAGCCACAC
2801 AGAAGGGCAA GAATCCCAAT TTAACTCATG TTATCATCAT TAGTGATCTG
2851 TGTTGTAGAA CATGAGGGTG TAAGCCTTCA GCCTGGCAAG TTACATGTAG
2901 AAAGCCCACA CTTGTGAAGG TTTTGTTTTA CAAATCACTT GATTTAACAC
2951 ACTCAGGTAG AATATTTTTA TTTTTACTGT TTTATACCCA GAAGTTATTT
3001 CTACATTGTT CTACAGCAAG AATATTCATA AAAGTATCCC TTTCAAATGC
3051 CTTTGAGAAG AATAGAAGAA AAAAAGTTTG TATATATTTT AAAAAATTGT
3101 TTTAAAAGTC AGTTTGCAAC ATGTCTGTAC CAAGATGGTA CTTTGCCTTA
3151 ACCGTTTATA TGCACTTTCA TGGAGACTGC AATACGTTGC TATGAGCACT
3201 TTCTTTATCC TTGGAGTTTA ATCCTTTGCT TCATCTTTCT ACAGTATGAC
3251 ATAATGATTT GCTATGTTGT AAAATCTTTG TAAAAAATTT CTATATAAAA
3301 ATATTTTGAA AATCTTAAAA AAAAAAAAAA AAA
```

BLAST Results

Entry HS389210 from database EMBL:
human STS SHGC-10164.
Score = 1592, P = 1.5e-64, identities = 346/364

Entry HS933343 from database EMBL:
human STS WI-16551.
Score = 1193, P = 5.7e-46, identities = 241/244

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 315 bp to 2027 bp; peptide length: 571
Category: strong similarity to known protein

```
  1 MTRAGDHNRQ RGCCGSLADY LTSAKFLLYL GHSLSTWGDR MWHFAVSVFL
 51 VELYGNSLLL TAVYGLVVAG SVLVLGAIIG DWVDKNARLK VAQTSLVVQN
101 VSVILCGIIL MMVFLHKHEL LTMYHGWVLT SCYILIITIA NIANLASTAT
151 AITIQRDWIV VVAGEDRSKL ANMNATIRRI DQLTNILAPM AVGQIMTFGS
201 PVIGCGFISG WNLVSMCVEY VLLWKVYQKT PALAVKAGLK EEETELKQLN
251 LHKDTEPKPL EGTHLMGVKD SNIHELEHEQ EPTCASQMAE PFRTFRDGWV
301 SYYNQPVFLA GMGLAFLYMT VLGFDCITTG YAYTQGLSGS ILSILMGASA
351 ITGIMGTVAF TWLRRKCGLV RTGLISGLAQ LSCLILCVIS VFMPGSPLDL
401 SVSPFEDIRS RFIQGESITP TKIPEITTEI YMSNGSNSAN IVPETSPESV
451 PIISVSLLFA GVIAARIGLW SFDLTVTQLL QENVIESERG IINGVQNSMN
501 YLLDLLHFIM VILAPNPEAF GLLVLISVSF VAMGHIMYFR FAQNTLGNKL
551 FACGPDAKEV RKENQANTSV V
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphute1_24j6, frame 3

TREMBLNEW:U76714_1 gene: "CAR1"; product: "cell adhesion regulator"; Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds., N = 1, Score = 1472, P = 7.2e-151

TREMBL:AC004683_5 gene: "T19C21.5"; Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence, complete sequence., N = 2, Score = 437, P = 2.8e-60

TREMBL:AF039046_2 gene: "R09B5.4"; Caenorhabditis elegans cosmid R09B5., N = 2, Score = 323, P = 1.5e-43

>TREMBLNEW:U76714_1 gene: "CAR1"; product: "cell adhesion regulator";
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds.
Length = 405

```
HSPs:

 Score = 1472 (220.9 bits), Expect = 7.2e-151, P = 7.2e-151
 Identities = 288/319 (90%), Positives = 297/319 (93%)

Query:     1 MTRAGDHNRQRGCCGSLADYLTSAKFLLYLGHSLSTWGDRMWHFAVSVFLVELYGNSLLL 60
             MT++ D    Q GCCGSLA+YLTSAKFLLYLGHSLSTWGDRMWHFAVSVFLVELYGNSLLL
Sbjct:     1 MTKSRDQTHQEGCCGSLANYLTSAKFLLYLGHSLSTWGDRMWHFAVSVFLVELYGNSLLL 60

Query:    61 TAVYGLVVAGSVLVLGAIIGDWVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKHEL 120
             TAVYGLVVAGSVLVLGAIIGDWVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHK+EL
Sbjct:    61 TAVYGLVVAGSVLVLGAIIGDWVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKNEL 120

Query:   121 LTMYHGWVLTSCYILIITIANIANLASTATAITIQRDWIVVVAGEDRSKLANMNATIRRI 180
             L MYHGWVLT CYILIITIANIANLASTATAITIQRDWIVVVAGE+RS+LA+MNATIRRI
Sbjct:   121 LNMYHGWVLTVCYILIITIANIANLASTATAITIQRDWIVVVAGENRSRLADMNATIRRI 180

Query:   181 DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK 240
             DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEY LLWKVYQKTPALAVKA LK
Sbjct:   181 DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYFLLWKVYQKTPALAVKAALK 240

Query:   241 EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGWV 300
              EE+ELKQL    KDTEPKPLEGTHLMG KDSNI ELE EQEPTCASQ+AEPFRTFRDGWV
Sbjct:   241 VEESELKQLTSPKDTEPKPLEGTHLMGEKDSNIRELECEQEPTCASQIAEPFRTFRDGWV 300

Query:   301 SYYNQPVFLAGMGLAF-LY 318
             SYYNQPVFL   G  F LY
Sbjct:   301 SYYNQPVFLGWHGPGFPLY 319
```

Pedant information for DKFZphutel_24j6, frame 3

Report for DKFZphutel_24j6.3

```
[LENGTH]        571
[MW]            62542.72
[pI]            6.08
[HOMOL]         TREMBL:U76714_1 gene: "CAR1"; product: "cell adhesion regulator"; Rattus
norvegicus cell adhesion regulator (CAR1) mRNA, complete cds. 1e-141
[BLOCKS]        BL00341D
[PROSITE]       MYRISTYL        15
[PROSITE]       MITOCH_CARRIER 1
[PROSITE]       CK2_PHOSPHO_SITE       6
[PROSITE]       PROKAR_LIPOPROTEIN     1
[PROSITE]       PKC_PHOSPHO_SITE       4
[PROSITE]       ASN_GLYCOSYLATION      4
[PFAM]          Laminin B (Domain IV)
[KW]            TRANSMEMBRANE  4
[KW]            LOW_COMPLEXITY      8.76 %

SEQ   MTRAGDHNRQRGCCGSLADYLTSAKFLLYLGHSLSTWGDRMWHFAVSVFLVELYGNSLLL
SEG   ............................................................
PRD   cccccccccccccccchhhhhhhheeeecccceeeccccchhhhhhhhheeeeecccccee
MEM   ..............................................MMMMMMMMMMMMM

SEQ   TAVYGLVVAGSVLVLGAIIGDWVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKHEL
SEG   .xxxxxxxxxxxxxxx............................................
PRD   ehhhhhhhcccceeeecccccccchhhhhhhhhhhhheeecccchhhhhhhhhhhhhhhhh
MEM   MMMMMMMMMMMMMMM...........MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ   LTMYHGWVLTSCYILIITIANIANLASTATAITIQRDWIVVVAGEDRSKLANMNATIRRI
SEG   .............xxxxxxxxxxxxxxxxxxxxx..........................
PRD   hhcccccchhhhhhhhhhhhhhhhhhhhhhheeeeccceeeeeeeccccchhhhhhhhhhh
MEM   MMMMMMM.....................................................

SEQ   DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK
SEG   ............................................................
PRD   hhhhhhccceeeceeeeeecceeeeeeeecccchhhhhhhhhhhhhhhccccchhhhhhhhh
MEM   ............................................................

SEQ   EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGWV
SEG   ............................................................
PRD   hhhhhhhhhhcccccccccccceeeeeeecccccccccccccccccccccccccccccccee
MEM   ............................................................

SEQ   SYYNQPVFLAGMGLAFLYMTVLGFDCITTGYAYTQGLSGSILSILMGASAITGIMGTVAF
SEG   ............................................................
PRD   eeecceeeecccchhhhhhcccccceeeeeeeeeccccceeeeeeeeccccceeeeehhhhhh
```

```
MEM    ...........................................................

SEQ    TWLRRKCGLVRTGLISGLAQLSCLILCVISVFMPGSPLDLSVSPFEDIRSRFIQGESITP
SEG    .........................................................xxx
PRD    hhhhhhccccccccchhhhhhhhhhhhhhhhcccccccccccccchhhhhhccccccccc
MEM    ...........................................................

SEQ    TKIPEITTEIYMSNGSNSANIVPETSPESVPIISVSLLFAGVIAARIGLWSFDLTVTQLL
SEG    xxxxxxxxxx..................................................
PRD    cccccccceeeeeccccccccccccccccccceeeeeehhhhhhhhhhccccchhhhhhhhhh
MEM    ..............................MMMMMMMMMMMMMMMMMMMMMMM......

SEQ    QENVIESERGIINGVQNSMNYLLDLLHFIMVILAPNPEAFGLLVLISVSFVAMGHIMYFR
SEG    ............................................................
PRD    hhhhhcccccceeeeccccchhhhhhhhhhheeeeecccccccceeeeeeeeccccccceeee
MEM    ............................MMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ    FAQNTLGNKLFACGPDAKEVRKENQANTSVV
SEG    ...............................
PRD    eeccccccceeeecccccchhhhhhhhcccccc
MEM    ...............................
```

Prosite for DKFZphutel_24j6.3

PS00001	100->104	ASN_GLYCOSYLATION	PDOC00001
PS00001	174->178	ASN_GLYCOSYLATION	PDOC00001
PS00001	434->438	ASN_GLYCOSYLATION	PDOC00001
PS00001	567->571	ASN_GLYCOSYLATION	PDOC00001
PS00005	23->26	PKC_PHOSPHO_SITE	PDOC00005
PS00005	176->179	PKC_PHOSPHO_SITE	PDOC00005
PS00005	294->297	PKC_PHOSPHO_SITE	PDOC00005
PS00005	487->490	PKC_PHOSPHO_SITE	PDOC00005
PS00006	16->20	CK2_PHOSPHO_SITE	PDOC00006
PS00006	36->40	CK2_PHOSPHO_SITE	PDOC00006
PS00006	294->298	CK2_PHOSPHO_SITE	PDOC00006
PS00006	396->400	CK2_PHOSPHO_SITE	PDOC00006
PS00006	403->407	CK2_PHOSPHO_SITE	PDOC00006
PS00006	445->449	CK2_PHOSPHO_SITE	PDOC00006
PS00008	12->18	MYRISTYL	PDOC00008
PS00008	65->71	MYRISTYL	PDOC00008
PS00008	76->82	MYRISTYL	PDOC00008
PS00008	193->199	MYRISTYL	PDOC00008
PS00008	267->273	MYRISTYL	PDOC00008
PS00008	311->317	MYRISTYL	PDOC00008
PS00008	336->342	MYRISTYL	PDOC00008
PS00008	339->345	MYRISTYL	PDOC00008
PS00008	353->359	MYRISTYL	PDOC00008
PS00008	368->374	MYRISTYL	PDOC00008
PS00008	373->379	MYRISTYL	PDOC00008
PS00008	435->441	MYRISTYL	PDOC00008
PS00008	461->467	MYRISTYL	PDOC00008
PS00008	490->496	MYRISTYL	PDOC00008
PS00008	494->500	MYRISTYL	PDOC00008
PS00013	122->133	PROKAR_LIPOPROTEIN	PDOC00013
PS00215	404->414	MITOCH_CARRIER	PDOC00189

Pfam for DKFZphutel_24j6.3

```
HMM_NAME      Laminin B (Domain IV)

HMM               *YWRlPERFLGDQvTsYGGkLe*
                   Y+R  +  LG+++ + G + +
Query         538  YFRFAQNTLGNKLFACGPDAK    558
```

DKFZphutel_2h3
----

group: differentiation/development

DKFZphutel_2h3 encodes a novel 267 amino acid protein, with similarity to ITM2 (integral membrane protein 2) of chicken and mouse.

The novel protein contains a prenyl group binding site (CAAX box) and seems to be post-translationally modified by the attachment of either a farnesyl or a geranyl-geranyl group. The similar gallus G. protein E25 a marker for chondro-osteogenic differentiation.

The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

strong similarity to mouse E25 and gallus E3-16

complete cDNA, EST hits
complete cds according to E25 start at Bp 56
putative transmembrane protein (1 TM)

Sequenced by AGOWA

Locus: unknown

Insert length: 2033 bp
Poly A stretch at pos. 2007, polyadenylation signal at pos. 1986

```
   1 GGACCGAGGC TGCACCGGCA GAGGCTGCGG GGCGGACGCG CGGGCCGGCG
  51 CAGCCATGGT GAAGATTAGC TTCCAGCCCG CCGTGGCTGG CATCAAGGGC
 101 GACAAGGCTG ACAAGGCGTC GGCGTCGGCC CCTGCGCCGG CCTCGGCCAC
 151 CGAGATCCTG CTGACGCCGG CTAGGGAGGA GCAGCCCCCA CAACATCGAT
 201 CCAAGAGGGG GAGCTCAGTG GGCGGCGTGT GCTACCTGTC GATGGGCATG
 251 GTCGTGCTGC TCATGGGCCT CGTGTTCGCC TCTGTCTACA TCTACAGATA
 301 CTTCTTTCTT GCACAGCTGG CCCGAGATAA CTTCTTCCGC TGTGGTGTGC
 351 TGTATGAGGA CTCCCTGTCC TCCCAGGTCC GGACTCAGAT GGAGCTGGAA
 401 GAGGATGTGA AAATCTACCT CGACGAGAAC TACGAGCGCA TCAACGTGCC
 451 TGTGCCCCAG TTTGGCGGCG GTGACCCTGC AGACATCATC CATGACTTCC
 501 AGCGGGGTCT GACTGCGTAC CATGATATCT CCCTGGACAA GTGCTATGTC
 551 ATCGAACTCA ACACCACCAT TGTGCTGCCC CCTCGCAACT TCTGGGAGCT
 601 CCTCATGAAC GTGAAGAGGG GGACCTACCT GCCGCAGACG TACATCATCC
 651 AGGAGGAGAT GGTGGTCACG GAGCATGTCA GTGACAAGGA GGCCCTGGGG
 701 TCCTTCATCT ACCACCTGTG CAACGGGAAA GACACCTACC GGCTCCGGCG
 751 CCGGGCAACG CGGAGGCGGA TCAACAAGCG TGGGGCCAAG AACTGCAATG
 801 CCATCCGCCA CTTCGAGAAC ACCTTCGTGG TGGAGACGCT CATCTGCGGG
 851 GTGGTGTGAG GCCCTCCTCC CCCAGAACCC CCTGCCGTGT TCCTCTTTTC
 901 TTCTTTCCAG CTGCTCTCTG GCCCTCCTCC TTCCCCCTGC TTAGCTTGTA
 951 CTTTGGACGC GTTTCTATAG AGGTGACATG TCTCTCCATT CCTCTCCAAC
1001 CCTGCCCACC TCCCTGTACC AGAGCTGTGA TCTCTCGGTG GGGGGCCCAT
1051 CTCTGCTGAC CTGGGTGTGG CGGAGGGAGA GGCGATGCTG CAAAGTGTTT
1101 TCTGTGTCCC ACTGTCTTGA AGCTGGGCCT GCCAAAGCCT GGGCCCACAG
1151 CTGCACCGGC AGCCCAAGGG GAAGGACCGG TTGGGGGAGC CGGGCATGTG
1201 AGGCCCTGGG CAAGGGGATG GGGCTGTGGG GGCGGGGCGG CATGGGCTTC
1251 AGAAGTATCT GCACAATTAG AAAAGTCCTC AGAAGCTTTT TCTTGGAGGG
1301 TACACTTTCT TCACTGTCCC TATTCCTAGA CCTGGGGCTT GAGCTGAGGA
1351 TGGGACGATG TGCCCAGGGA GGGACCCACC AGAGCACAAG AGAAGGTGGC
1401 TACCTGGGGG TGTCCCAGGG ACTCTGTCAG TGCCTTCAGC CCACCAGCAG
1451 GAGCTTGGAG TTTGGGGAGT GGGGATGAGT CCGTCAAGCA CAACTGTTCT
1501 CTGAGTGGAA CCAAAGAAGC AAGGAGCTAG GACCCCCAGT CCTGCCCCCC
1551 AGGAGCACAA GCAGGGTCCC CTCAGTCAAG GCAGTGGGAT GGGCGGCTGA
1601 GGAACGGGGC AGGCAAGGTC ACTGCTCAGT CACGTCCACG GGGGACGAGC
1651 CGTGGGTTCT GCTGAGTAGG TGGAGCTCAT TGCTTTCTCC AAGCTTGGAA
1701 CTGTTTTGAA AGATAACACA GAGGGAAAGG GAGAGCCACC TGGTACTTGT
1751 CCACCCTGCC TCCTCTGTTC TGAAATTCCA TCCCCCTCAG CTTAGGGGAA
1801 TGCACCTTTT TCCCTTTCCT TCTCACTTTT GCATGTTTTT ACTGATCATT
1851 CGATATGCTA ACCGTTCTCA GCCCTGAGCC TTGGAGAGGA GGGCTGTAAC
1901 GCCTTCAGTC AGTCTCTGGG GATGAAACTC TTAAATGCTT TGTATATTTT
1951 CTCAATTAGA TCTCTTTTCA GAAGTGTCTA TAGAACAATA AAAATCTTTT
2001 ACTTCTGAAA AAAAAAAAAA AAAAGGGCGG CCG
```

BLAST Results

Entry B64417 from database EMBL:
CIT-HSP-2023A7.TR CIT-HSP Homo sapiens genomic clone 2023A7.
Length = 715
Plus Strand HSPs:

```
Score = 1546 (232.0 bits), Expect = 7.8e-64, P = 7.8e-64
Identities = 310/311 (99%)
```

Medline entries

96325063:
Isolation of markers for chondro-osteogenic differentiation using cDNA library subtraction.
Molecular cloning and characterization of a gene belonging to a novel multigene family of
integral membrane proteins.

Peptide information for frame 2

ORF from 56 bp to 856 bp; peptide length: 267
Category: strong similarity to known protein

```
  1 MVKISFQPAV AGIKGDKADK ASASAPAPAS ATEILLTPAR EEQPPQHRSK
 51 RGSSVGGVCY LSMGMVVLLM GLVFASVYIY RYFFLAQLAR DNFFRCGVLY
101 EDSLSSQVRT QMELEEDVKI YLDENYERIN VPVPQFGGGD PADIIHDFQR
151 GLTAYHDISL DKCYVIELNT TIVLPPRNFW ELLMNVKRGT YLPQTYIIQE
201 EMVVTEHVSD KEALGSFIYH LCNGKDTYRL RRRATRRRIN KRGAKNCNAI
251 RHFENTFVVE TLICGVV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel_2h3, frame 2

SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16)., N = 1, Score = 573, P = 1.3e-55

SWISSNEW:ITMB_MOUSE INTEGRAL MEMBRANE PROTEIN 2B (E25B PROTEIN)., N = 1, Score = 560, P = 3.2e-54

SWISSNEW:ITMA_HUMAN INTEGRAL MEMBRANE PROTEIN 2A (E25 PROTEIN)., N = 1, Score = 456, P = 3.3e-43

>SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN
E3-16).
Length = 262

HSPs:

```
Score = 573 (86.0 bits), Expect = 1.3e-55, P = 1.3e-55
Identities = 117/264 (44%), Positives = 172/264 (65%)

Query:    1 MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY 60
            MVK+SF  A+A    + A+K   ++         ++L+ P   ++P        G
Sbjct:    1 MVKVSFNSALA--HKEAANKEEENS-------QVLILPPDAKEPEDVVVPAGHKRAWCWC 51

Query:   61 LSMGMVVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLY-EDSLS-----SQVRTQM-- 112
            +  G+  +L G++     Y+Y+YF   Q        + CG+ Y ED LS       +Q+++
Sbjct:   52 MCFGLAFMLAGVILGGAYLYKYFAFQQ---GGVYFCGIKYIEDGLSLPESGAQLKSARYH 108

Query:  113 ELEEDVKIYLDENYERINVPVPQFGGGDPADIIHDFQRGLTAYHDISLDKCYVIELNTTI 172
             +E++++I  +E+ E I+VPVP+F   DPADI+HDF R LTAY D+SLDKCYVI LNT++
Sbjct:  109 TIEQNIQILEEEDVEFISVPVPEFADSDPADIVHDFHRRLTAYLDLSLDKCYVIPLNTSV 168

Query:  173 VLPPRNFWELLMNVKRGTYLPQTYIIQEEMVVTEHVSDKEALGSFIYHLCNGKDTYRLRR 232
            V+PP+NF ELL+N+K GTYLPQ+Y+I E+M+VT+ + + + LG FIY LC GK+TY+L+R
Sbjct:  169 VMPPKNFLELLINIKAGTYLPQSYLIHEQMIVTDRIENVDQLGFFIYRLCRGKETYKLQR 228

Query:  233 RATRRRINKRGAKNCNAIRHFENTFVVETLIC 264
            +    + I KR A NC  IRHFEN F +ETLIC
Sbjct:  229 KEAMKGIQKREAVNCRKIRHFENRFAMETLIC 260
```

Pedant information for DKFZphutel_2h3, frame 2

Report for DKFZphutel_2h3.2

```
[LENGTH]      267
[MW]          30253.96
[pI]          8.16
[HOMOL]       SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).
1e-49
[PROSITE]     MYRISTYL       4
[PROSITE]     PRENYLATION    1
[PROSITE]     CAMP_PHOSPHO_SITE     3
[PROSITE]     CK2_PHOSPHO_SITE      3
[PROSITE]     TYR_PHOSPHO_SITE      1
[PROSITE]     PKC_PHOSPHO_SITE      4
[PROSITE]     ASN_GLYCOSYLATION     1
[KW]          TRANSMEMBRANE   1
[KW]          LOW_COMPLEXITY     15.36 %

SEQ    MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY
SEG    ..............xxxxxxxxxxxxxxx.............................
PRD    ccccccccchhhhhhhhhhhhhhhhhcccccccceeeccccccccccccccccccccchh
MEM    ........................................................MMMM

SEQ    LSMGMVVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLYEDSLSSQVRTQMELEEDVKI
SEG    ..xxxxxxxxxxx...............................................
PRD    hhhhhhhhhhhhhhhhhhcchhhhhhhhhhccceeeeeeccccccccchhhhhhhhhhhh
MEM    MMMMMMMMMMMMMMMMMMMMMMMM....................................

SEQ    YLDENYERINVPVPQFGGGDPADIIHDFQRGLTAYHDISLDKCYVIELNTTIVLPPRNFW
SEG    ............................................................
PRD    hhccccceeeeccccccccccccchhhhhhhhhhhhhhhccccceeeeeccceeeccccchhh
MEM    ............................................................

SEQ    ELLMNVKRGTYLPQTYIIQEEMVVTEHVSDKEALGSFIYHLCNGKDTYRLRRRATRRRIN
SEG    ...............................................xxxxxxxxxxx
PRD    hhhhhhccccccceeeeehhhhhhhccccchhhhhheeecccccchhhhhhhhhhhhhh
MEM    ............................................................

SEQ    KRGAKNCNAIRHFENTFVVETLICGVV
SEG    xx.........................
PRD    hhhhccceeeeccccchhhhhheeeccc
MEM    ...........................
```

Prosite for DKFZphutel_2h3.2

PS00001	169->173	ASN_GLYCOSYLATION	PDOC00001
PS00004	50->54	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	187->191	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	227->230	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00007	119->127	TYR_PHOSPHO_SITE	PDOC00007
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	243->249	MYRISTYL	PDOC00008
PS00294	264->268	PRENYLATION	PDOC00266

(No Pfam data available for DKFZphutel_2h3.2)

DKFZphmcf1_1a11
----

group: transmembrane protein

DKFZphmcf1_1a11 encodes a novel 393 amino acid protein with weak similarity to S.pombe SPBC29A3_3 protein and S. cerevisiae putative membrane protein YDR255c.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes and as a new marker for mammary carcinoma cells.

similarity to YDR255c and SPBC29A3.03c
membrane regions: 1
Summary DKFZphmcf1_1a11 encodes a novel 393 amino acid protein, with similarity to YDR255c and SPBC29A3.03c.

similarity to YDR255c and SPBC29A3.03c

complete cDNA, complete cds, EST hits
potential start at Bp 110 matches kozak consensus

Sequenced by DKFZ

Locus: /map="542.7 cR from top of Chr5 linkage group"

Insert length: 1819 bp
Poly A stretch at pos. 1808, no polyadenylation signal found

```
   1 CCCGGCCCAG CCCCCGAAGA GCCGCCTCAG CCGGGGGGAG TTGCTCGGAC
  51 TCAAACGTCC AGTCCTCGTG CGACCGCGCT GGGTCGGAAG TGAGCAGGCT
 101 GAGGCCACCA TGGAGCAGTG TGCGTGCGTG GAGAGAGAGC TGGACAAGGT
 151 CCTGCAGAAG TTCCTGACCT ACGGGCAGCA CTGTGAGCGG AGCCTGGAGG
 201 AGCTGCTGCA CTACGTGGGC CAGCTGCGGG CTGAGCTGGC CAGCGCAGCC
 251 CTCCAGGGGA CCCCTCTCTC AGCCACCCTC TCTCTGGTGA TGTCACAGTG
 301 CTGCCGGAAG ATCAAAGATA CGGTGCAGAA ACTGGCTTCG GACCATAAGG
 351 ACATTCACAG CAGTGTATCC CGAGTGGGCA AAGCCATTGA CAGGAACTTC
 401 GACTCTGAGA TCTGTGGTGT TGTGTCAGAT GCGGTGTGGG ACGCGCGGGA
 451 ACAGCAGCAG CAGATCCTGC AGATGGCCAT CGTGGAACAC CTGTATCAGC
 501 AGGGCATGCT CAGCGTGGCC GAGGAGCTGT GCCAGGAATC AACGCTGAAT
 551 GTGGACTTGG ATTTCAAGCA GCCTTTCCTA GAGTTGAATC GAATCCTGGA
 601 AGCCCTGCAC GAACAAGACC TGGGTCCTGC GTTGGAATGG GCCGTCTCCC
 651 ACAGGCAGCG CCTGCTGGAA CTCAACAGCT CCCTGGAGTT CAAGCTGCAC
 701 CGACTGCACT TCATCCGCCT CTTGGCAGGA GGCCCCGCGA AGCAGCTGGA
 751 GGCCCTCAGC TATGCTCGGC ACTTCCAGCC CTTTGCTCGG CTGCACCAGC
 801 GGGAGATCCA GGTGATGATG GGCAGCCTGG TGTACCTGCG GCTGGGCTTG
 851 GAGAAGTCAC CCTACTGCCA CCTGCTGGAC AGCAGCCACT GGGCAGAGAT
 901 CTGTGAGACC TTTACCCGGG ACGCCTGTTC CCTGCTGGGG CTTTCTGTGG
 951 AGTCCCCCCT TAGCGTCAGC TTTGCCTCTG GCTGTGTGGC GCTGCCTGTG
1001 TTGATGAACA TCAAGGCTGT GATTGAGCAG CGGCAGTGCA CTGGGGTCTG
1051 GAATCACAAG GACGAGTTAC CGATTGAGAT TGAACTAGGC ATGAAGTGCT
1101 GGTACCACTC CGTGTTCGCT TGCCCCATCC TCCGCCAGCA GACGTCAGAT
1151 TCCAACCCTC CCATCAAGCT CATCTGTGGC CATGTTATCT CCCGAGATGC
1201 ACTCAATAAG CTCATTAATG GAGGAAAGCT GAAGTGTCCC TACTGTCCCA
1251 TGGAGCAGAA CCCGGCAGAT GGGAAACGCA TCATATTCTG ATTCCTACCT
1301 GGAAGGAATT TTGTTGAAAG GGGTTTTCAC CTGTGAGCCT TGGTCTGTCT
1351 CGGTAGGGTG GTCAACTTCA GTGGACTGTG GTTGGTTTCA GAGCGCCTGG
1401 CTGAGGAGTT CCACTGAGGG GAGCACTGGA GCAGCCCTTT GGCAGAGGCT
1451 GAGGAGGGAG ATGGACCAGC CCACGCCTGG CACCTGGCTC CATGGCATAA
1501 GGAAAGGGAG ATGCTGGCCT CTGTGCTCCT GCTGTCTTTT CCTGTTTCTG
1551 TTTGCGTTTG ACTTAGTAGC AACCGACAGA GTGGCAAGGG ATTTGGTCTT
1601 CAGCAGTAGA CATCCTTCCA CCCCTGCCCT CAGCCAAGTC TCTTGCTGCC
1651 ATGCCAATGC TATGTCCACC CTTGCCCCTC GGCCCAAGAG TGTCCAGCGG
1701 TGGCCCACCT CTTCCTCCCA CTACAGCCTC AACAGTATGT ACCATCTCCC
1751 ACTGTAAATA GTCCCAGTTA GAACGGAATG CCGTTGTTTT ATAACTTTGA
1801 ACAAATGTAA AAAAAAAAA
```

BLAST Results

Entry HS579359 from database EMBL:
human STS WI-6350.
Score = 1027, P = 9.9e-40, identities = 207/209

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 110 bp to 1288 bp; peptide length: 393
Category: similarity to unknown protein

```
  1 MEQCACVERE LDKVLQKFLT YGQHCERSLE ELLHYVGQLR AELASAALQG
 51 TPLSATLSLV MSQCCRKIKD TVQKLASDHK DIHSSVSRVG KAIDRNFDSE
101 ICGVVSDAVW DAREQQQQIL QMAIVEHLYQ QGMLSVAEEL CQESTLNVDL
151 DFKQPFLELN RILEALHEQD LGPALEWAVS HRQRLLELNS SLEFKLHRLH
201 FIRLLAGGPA KQLEALSYAR HFQPFARLHQ REIQVMMGSL VYLRLGLEKS
251 PYCHLLDSSH WAEICETFTR DACSLLGLSV ESPLSVSFAS GCVALPVLMN
301 IKAVIEQRQC TGVWNHKDEL PIEIELGMKC WYHSVFACPI LRQQTSDSNP
351 PIKLICGHVI SRDALNKLIN GGKLKCPYCP MEQNPADGKR IIF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_1a11, frame 2

TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3., N = 2, Score = 302, P = 3.4e-42

PIR:S67312 probable membrane protein YDR255c - yeast (Saccharomyces cerevisiae), N = 1, Score = 271, P = 5.3e-22

TREMBL:CET07D1_2 gene: "T07D1.2"; Caenorhabditis elegans cosmid T07D1., N = 1, Score = 193, P = 5.6e-13

```
>TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein";
     S.pombe chromosome II cosmid c29A3.
           Length = 398

  HSPs:

 Score = 302 (45.3 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42
 Identities = 55/142 (38%), Positives = 89/142 (62%)

Query:   252 YCHLLDSSHWAEICETFTRDACSLLGLSVESPLSVSFASGCVALPVLMNIKAVIEQRQCT 311
             Y  +LD   W  +   F R+ C+ LG+S+ESPL +   +G +ALP+L++ + ++++++
Sbjct:   258 YIDVLDLD-WKSLELLFVREFCAALGMSLESPLDIVVNAGAIALPILLKMSSIMKKKHTE 316

Query:   312 GVWNHKDELPIEIELGMKCWYHSVFACPILRQQTSDSNPPIKLICGHVISRDALNKLING 371
               W  + ELP+EI L     +HSVF CP+ ++Q ++ NPP+ + CGHVI +++L +L
Sbjct:   317 --WTSQGELPVEIFLPSSYHFHSVFTCPVSKEQATEENPPMMMSCGHVIVKESLRQLSRN 374

Query:   372 G--KLKCPYCPMEQNPADGKRIIF 393
             G  + KCPYCP E   AD  R+ F
Sbjct:   375 GSQRFKCPYCPNENVAADAIRVYF 398

 Score = 161 (24.2 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42
 Identities = 51/221 (23%), Positives = 102/221 (46%)

Query:    22 GQHCERSLEELLHYVGQLRAELASAALQGTPLSATLSLVMSQCCRKIKDTVQKLASDHKD 81
             G C   L EL       + +  + L+  P ++    LV   C  K +      L    K
Sbjct:    15 GNKCLAKLNEL----ESILKDAKKSCLKD-PTTSMKELVA--CSEKTQQVFDDLKRTEKK 67

Query:    82 IHSSVSRVGKAIDRNFDSEICGVVSDAVWDAREQQQQILQMAIVEHLYQQGMLSVAEELC 141
             H+S++R GK +++ F+ ++   +   + +++++++   +  A+  H ++QG + +A   C
Sbjct:    68 FHTSLNRFGKTLEKKFNFDLEDIKLHSSFESKKRE---IDTALSLHFFRQGDVELAHLFC 124

Query:   142 QESTLNVDLDFKQPFLELNRILEALHEQDLGPALEWAVSHRQRLLELNSSLEFKLHRLHF 201
             +E+ +     +     F  L  I++ + ++DL    +EWA   R  L   SSLE+ L +
Sbjct:   125 KEAGIEEPSESLHVFTLLKSIVQGIRDKDLKLPIEWASQCRGYLERKGSSLEYTLQKYRL 184

Query:   202 IRLLAGGPAKQL-EALSYAR-HFQPFARLHQREIQVMMGSLVY 242
             +        K +  A+ Y R +   F + H  +IQ  M +L +
```

Sbjct: 185 VSNYL--TTKDIMAAIRYCRTNMAEFQKKHLADIQKTMIALFF 225

Pedant information for DKFZphmcf1_1a11, frame 2

Report for DKFZphmcf1_1a11.2

[LENGTH] 393
[MW] 44414.77
[pI] 6.15
[HOMOL] TREMBL:SPBC29A3_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3. 2e-39
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YDR255c] 8e-23
[PIRKW] transmembrane protein 2e-21
[PROSITE] MYRISTYL 2
[PROSITE] AMIDATION 1
[PROSITE] CK2_PHOSPHO_SITE 3
[PROSITE] PROKAR_LIPOPROTEIN 1
[PROSITE] TYR_PHOSPHO_SITE 3
[PROSITE] PKC_PHOSPHO_SITE 1
[PROSITE] ASN_GLYCOSYLATION 1
[KW] TRANSMEMBRANE 1

```
SEQ MEQCACVERELDKVLQKFLTYGQHCERSLEELLHYVGQLRAELASAALQGTPLSATLSLV
PRD ccceeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccchhhhh
MEM ............................................................

SEQ MSQCCRKIKDTVQKLASDHKDIHSSVSRVGKAIDRNFDSEICGVVSDAVWDAREQQQQIL
PRD hhhhhhhhhhhhhhhhhhhhhcccccccchhhhhccccceeeechhhhhhhhhhhhhhh
MEM ............................................................

SEQ QMAIVEHLYQQGMLSVAEELCQESTLNVDLDFKQPFLELNRILEALHEQDLGPALEWAVS
PRD hhhhhhhhhhhcccchhhhhhhhhhhcccccccccchhhhhhhhhhhhhhcccccchhhhhhh
MEM ............................................................

SEQ HRQRLLELNSSLEFKLHRLHFIRLLAGGPAKQLEALSYARHFQPFARLHQREIQVMMGSL
PRD hhhhhhhccccchhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM ............................................................

SEQ VYLRLGLEKSPYCHLLDSSHWAEICETFTRDACSLLGLSVESPLSVSFASGCVALPVLMN
PRD hhcccccccccccccccchhhhhhhhhhhhhhhhhhhhccccccceeeeccccccchhhhhh
MEM ....................................MMMMMMMMMMMMMMMMMMMMMMMM

SEQ IKAVIEQRQCTGVWNHKDELPIEIELGMKCWYHSVFACPILRQQTSDSNPPIKLICGHVI
PRD hhhhhhhhhhhcccccccccceeeeeecccceeeeeeeeecchhhhhccccccccccccccceee
MEM MMMMMM......................................................

SEQ SRDALNKLINGGKLKCPYCPMEQNPADGKRIIF
PRD eehhhhhhhcccccccccccccccchhhhhcccc
MEM .................................
```

Prosite for DKFZphmcf1_1a11.2

PS00001	189->193	ASN_GLYCOSYLATION	PDOC00001
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	27->36	TYR_PHOSPHO_SITE	PDOC00007
PS00007	244->253	TYR_PHOSPHO_SITE	PDOC00007
PS00008	37->43	MYRISTYL	PDOC00008
PS00008	50->56	MYRISTYL	PDOC00008
PS00009	387->391	AMIDATION	PDOC00009
PS00013	282->293	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphmcf1_1a11.2)

DKFZphmcf1_1c23
----

group: mammary carcinoma derived

DKFZphmcf1_1c23.1 encodes a novel 311 amino acid proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of mamma carcinoma-specific genes.

unknown, proline rich protein

complete cDNA, complete cds? potential start at Bp 50, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3077 bp
Poly A stretch at pos. 3067, polyadenylation signal at pos. 3048

```
   1 AACTGGCCCC CTCCCCCACC CCCTGCCCCT GAGGAGCAGG ACCTGTCCAT
  51 GGCTGACTTC CCCCCACCAG AGGAGGCTTT TTTCTCTGTG GCCAGCCCTG
 101 AGCCTGCAGG CCCTTCAGGC TCCCCAGAGC TTGTCAGCTC CCCGGCTGCT
 151 TCGTCCTCCT CAGCTACTGC TTTGCAGATT CAGCCCCCGG GTAGCCCAGA
 201 CCCTCCTCCA GCTCCGCCAG CCCCAGCTCC TGCTAGTTCC GCCCCAGGGC
 251 ATGTGGCCAA GCTCCCTCAG AAGGAACCGG TGGGCTGTAG CAAGGGTGGT
 301 GGGCCTCCCA GGGAGGACGT AGGTGCGCCC CTGGTCACGC CCTCGCTCCT
 351 GCAGATGGTG CGGCTGCGCT CCGTGGGTGC TCCAGGAGGG GCTCCCACCC
 401 CAGCACTGGG GCCATCGGCC CCCCAGAAAC CACTGCGAAG GGCCCTGTCA
 451 GGGCGGGCCA GCCCAGTGCC TGCCCCCTCC TCAGGGCTCC ATGCTGCGGT
 501 CCGACTCAAG GCCTGCAGCC TGGCCGCCAG TGAAGGCCTC TCAAGTGCTC
 551 AGCCCAACGG ACCGCCTGAG GCAGAGCCAC GGCCTCCCCA GTCCCCTGCC
 601 TCAACGGCCA GTTTCATCTT CTCCAAGGGC TCTAGGAAGC TGCAGCTGGA
 651 GCGGCCCGTG TCCCCTGAGA CCCAGGCTGA CCTCCAGCGG AATCTGGTGG
 701 CAGAACTCCG GAGCATCTCA GAGCAGCGGC CACCCCAGGC CCCAAAGAAG
 751 TCACCTAAGG CTCCCCCACC TGTGGCCCGC AAGCCGTCTG TGGGAGTCCC
 801 CCCACCCGCC TCCCCCAGTT ACCCTCGAGC TGAGCCCCTT ACTGCTCCTC
 851 CCACCAATGG GCTCCCTCAC ACCCAGGACA GGACTAAGAG GGAGCTGGCG
 901 GAGAATGGAG GTGTCCTGCA GCTGGTGGGC CCAGAGGAGA AGATGGGCCT
 951 CCCGGGCTCA GACTCACAGA AAGAGCTGGC CTGACCACCA GGCACCTCAC
1001 TGGCACTGCT GACCCATCCC AGAAACACAA TCTCAGGGAC CCGAGCAGCT
1051 CCAAGGACGA GAGGATACAG CAGACACAAC CTAATAGAGA GGGCGCCTGC
1101 AGCCTTAACC TCCACGGCCT TCGATACTTA TGCAAGCCTG GTGTTGCTCC
1151 TGTCCTCAGA GTCATCCTGC GCTCATGCCT TTTCCCGAAT GGGTTCACCT
1201 CTGGCAGTTG CCGCTTCAGT CTTGGCCTTA GCCTCATCTT GAAGTGGGTA
1251 GCTGGCGGGA GAGGGTGGCT GCGCCCCCTG CTGGCCCTGA GGCTGCAGAG
1301 TTGGGAGCAG GACACCTCAC CTGAGTTTCA TTTTTTTTCA TGTCCAAACC
1351 ATGCACATAC TATAGTCCAG AATCAAAGCA CTTTTGAAAA GTGGCTGCAT
1401 GGCCATCCTC CAGGGCCCAG GAAGTTGCAT TCCAAGGGCC TGTTTACATG
1451 GCAGCAGAAT CCATCCCCGG CAGTCAGCCC ATAGCTTGGG ACCAGTCTGT
1501 GCCCTCCTGC CCAGTCCAGT TTACTCCTCT TGGTTCCTGA AGGTGGCCAA
1551 GTCATTGTGT TCCCACAGGC TTCTCTAGGC TGGGGGCAGG TGTGGGGCTG
1601 TGGAATTCCA AAGCACAAAA GGTGCAGAGG GGATTGGCCT TCCTGTGCCT
1651 CAACTCACCA ACCACCCTCC TGCCTTCCAG TTCTGCCAGG TGCTCCATGC
1701 TGGGGACAAG TAGGAGACTG CCAGGGCCCA AAGAAATGGG TGAGCAGTAG
1751 AGTCATCTCG GGGCACTTGG CAGTGTCAAG CACCTGCCCC TTGCCTCCTT
1801 GACCACACTG GGGTGGGTGG GCCCCCAGCA CTTCAGAGGC AGGAGCCTTT
1851 GGGCTGAGCA AGCACTGAGG AGGTGGATGG AAGGGAGCAT CTGGAGGGGG
1901 GGAGCTTCCT TGAGCAGTGG GCCCAGGCCT GGCCCTCCAC ACTTCATTCT
1951 CTGACCTTTC TCTCTCCTCA TTTCGGTGCA TGTCCTTTCT GCAGCTGCCT
2001 TTCAGCACAG GTGGTTCCAC TGGGGGCAGC TAACGCTGAG TGACAAGGAT
2051 GGGAAGCCAC AGGTGCATTT TACTCAAGTC TTCTCTAGTC AATGAGGGGC
2101 ACCCAGTGCT TCTAGGGCAG GCTGGGTGGT GGTCCCCTAG GTATCAGCCT
2151 CTCTTACTGT ACTCTCCGGG AATGTTAACC TTTCTATTTT CAGCCTGTGC
2201 CACCTGTCTA GGCAAGCTGG CTTCCCCATT GGCCCCTGTG GGTCCACAGC
2251 AGCGTGGCTG CCCCCCAGGG CCACCGCTTC TTTCTTGATC CTCTTTCCTT
2301 AACAGTGACT TGGGCTTGAG TCTGGCAAGG AACCTTGCTT TTAGCTTCAC
2351 CACCAAGGAG AGAGGTTGAC ATGACCTCCC CGCCCCCTCA CCAAGGCTGG
2401 GAACAGAGGG GATGTGGTGA GAGCCAGGTT CCTCTGGCCC TCTCCAGGGT
2451 GTTTTCCACT AGTCACTACT GTCTTCTCCT TGTAGCTAAT CAATCAATAT
2501 TCTTCCCTTG CCTGTGGGCA GTGGAGAGTG CTGCTGGGTG TACGCTGCAC
2551 CTGCCCACTG AGTTGGGGAA AGAGGATAAT CAGTGAGCAC TGTTCTGCTC
2601 AGAGCTCCTG ATCTACCCCA CCCCCTAGGA TCCAGGACTG GGTCAAAGCT
2651 GCATGAAACC AGGCCCTGGC AGCAACCTGG GAATGGCTGG AGGTGGGAGA
2701 GAACCTGACT TCTCTTTCCC TCTCCCTCCT CCAACATTAC TGGAACTCTA
```

```
2751 TCCTGTTAGG ATCTTCTGAG CTTGTTTCCC TGCTGGGTGG GACAGAGGAC
2801 AAAGGAGAAG GGAGGGTCTA GAAGAGGCAG CCCTTCTTTG TCCTCTGGGG
2851 TAAATGAGCT TGACCTAGAG TAAATGGAGA GACCAAAAGC CTCTGATTTT
2901 TAATTTCCAT AAAATGTTAG AAGTATATAT ATACATATAT ATATTTCTTT
2951 AAATTTTTGA GTCTTTGATA TGTCTAAAAA TCCATTCCCT CTGCCCTGAA
3001 GCCTGAGTGA GACACATGAA GAAAACTGTG TTTCATTTAA AGATGTTAAT
3051 TAAATGATTG AAACTTGAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 49 bp to 981 bp; peptide length: 311
Category: putative protein
Classification: unset

```
  1 MADFPPPEEA FFSVASPEPA GPSGSPELVS SPAASSSSAT ALQIQPPGSP
 51 DPPPAPPAPA PASSAPGHVA KLPQKEPVGC SKGGGPPRED VGAPLVTPSL
101 LQMVRLRSVG APGGAPTPAL GPSAPQKPLR RALSGRASPV PAPSSGLHAA
151 VRLKACSLAA SEGLSSAQPN GPPEAEPRPP QSPASTASFI FSKGSRKLQL
201 ERPVSPETQA DLQRNLVAEL RSISEQRPPQ APKKSPKAPP PVARKPSVGV
251 PPPASPSYPR AEPLTAPPTN GLPHTQDRTK RELAENGGVL QLVGPEEKMG
301 LPGSDSQKEL A
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_1c23, frame 1

PIR:S49915 extensin-like protein - maize, N = 1, Score = 215, P = 6.1e-15

PIR:A28996 proline-rich protein M14 precursor - mouse, N = 1, Score = 191, P = 3.8e-13

>PIR:S49915 extensin-like protein - maize
Length = 1,188

HSPs:

Score = 215 (32.3 bits), Expect = 6.1e-15, P = 6.1e-15
Identities = 81/269 (30%), Positives = 115/269 (42%)

```
Query:      5 PPPEEAFFS----VASPEPAGPSGSPELVSSPAASSSSATALQIQPPGSP--DPPP---A 55
              PPP     S    V SP P  P  SP     PA  +SS   ++  PP +P   PPP   +
Sbjct:    598 PPPPAPVASPPPPVKSPPPPTPVASPP---PPAPVASSPPPMKSPPPPTPVSSPPPPEKS 654

Query:     56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGA 115
              PP P PA S P    + P   P    K   PP + +  P + PS     +      P
Sbjct:    655 PPPPPPAKSTPPP-EEYPT--PPTSVKSSPPPEKSLPPPTLIPSPPPQEKPTPPSTPSKP 711

Query:    116 PTPALGPSAPQKPLRRA-LSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174
              P+    PS P++P+     + ++SP PAP S           +LA      S + + PP
Sbjct:    712 PSSPEKPSPPKEPVSSPPQTPKSSPPPAPVSSPPPTPVSSPPALAPVSSPPSVKSSPPPA 771

Query:    175 AEPRPPQSPASTASFIFSKGSRKLQLERPV-SPETQADLQRNLVAELRSISEQRPPQAPK 233
                  PP +P   +S         +Q+  P  +P++   L    V+    + +  PP AP
Sbjct:    772 PLSSPPPAPQVKSS------PPPVQVSSPPPAPKSSPPLAP--VSSPPQVEKTSPPPAPL 823

Query:    234 KSPKAPPPVARKPSVGV--PPPASPSYPRAEPLTAPPTNGLP 273
               SP   P  +  P V V  PPP   S P   P+++PP    P
Sbjct:    824 SSPPLAPK-SSPPHVVVSSPPPVVKSSPPPAPVSSPPLTPKP 864
```

Score = 206 (30.9 bits), Expect = 9.1e-14, P = 9.1e-14

```
 Identities = 82/261 (31%), Positives = 108/261 (41%)

Query:    17 PEPAG-PSGSPELVSSPAASS---SSATALQIQPPGSPDPPPAP---PAPAPASSAPGHV 69
             P P G P  SP +    PAAS+   S  T   + P   P+P P P   P P P   +P
Sbjct:   410 PTPGGGPPSSP-VPGKPAASAPMPSPHTPPDVSPEPLPEPSPVPAPAPMPMPTPHSPPAD 468

Query:    70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTPALGPSAPQKP 128
              +P    PV G S       P    V   P  +      +V+L    AP G+P P +  ++P  P
Sbjct:   469 DYVPPTPPVPGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTSPPAP 528

Query:   129 LRRALSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSPASTAS 188
             +      G  SP P P S +      +K+    A   G    + P  PPE  P PP    AS
Sbjct:   529 I-----GSPSP-PPPVSVVSPPPPVKSPPPPAPVG---SPP--PPEKSPPPPAPVASPPP 577

Query:   189 FIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKSPKAPPPVARKPS- 247
              + S      L     P  P  ++       VA        + PP  P  SP  P PVA  P
Sbjct:   578 PVKSPPPPTLVASPP--PPVKSPPPPAPVASPPPPVKSPPPPTPVASPPPPAPVASSPPP 635

Query:   248 VGVPPP----ASPSYPRAEPLTAPPTNGLPHTQD 277
             +  PPP    +SP  P   P   PP   P  ++
Sbjct:   636 MKSPPPPTPVSSPPPPEKSPPPPPPAKSTPPPEE 669

 Score = 202 (30.3 bits), Expect = 2.9e-13, P = 2.9e-13
 Identities = 81/254 (31%), Positives = 110/254 (43%)

Query:    16 SPEPAGPSGSPELV--SSP--AASSSSATALQIQPPGSP-DPPPAPPAPAPASSAPGHVA 70
             SP PA P  SP L   SSP      SS       ++  PP +P    PP  P PA   S P HV+
Sbjct:   817 SPPPA-PLSSPPLAPKSSPPHVVVSSPPPVVKSSPPPAPVSSPPLTPKPA---SPPAHVS 872

Query:    71 KLPQ----KEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTPALGPSAPQ 126
               P+       P    +     PP E    +P  TP  L      ++S  P     +P + P +
Sbjct:   873 SPPEVVKPSTPPAPTTVISPPSEPKSSPPPTPVSLPPPIVKSSPPPAMVSSPPMTPKSSP 932

Query:   127 KPLRRAL---SGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSP 183
              P+  +     + ++SP PAP S    A    K+    A    L  P  PPE +  PP +P
Sbjct:   933 PPVVVSSPPPTVKSSPPPAPVSSPPATP--KSSPPPAPVNL----P--PPEVKSSPPPTP 984

Query:   184 ASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKSPKAPPPVA 243
              S+               +  P  PE ++       V+       +  PP AP   SP   PPPV
Sbjct:   985 VSSPPPAPKSSPPPAPMSSPPPPEVKSPPPPAPVSSPPPPVKSPPPPAPVSSP--PPPVK 1042

Query:   244 RKPS---VGVPPPASPSYPRAEPLTAPP 268
               P     V  PPP   S P    P++++PP
Sbjct:  1043 SPPPPAPVSSPPPPVKSPPPPAPISSPP 1070

 Score = 190 (28.5 bits), Expect = 7.9e-12, P = 7.9e-12
 Identities = 74/264 (28%), Positives = 111/264 (42%)

Query:     5 PPPEEAFFSVASPEPAGPSGSPELVSSPAAS-SSSATALQIQPPGSPDPPPAPPAPAPAS 63
             PPP       S    PE + P    P    + P       +  T+++  PP      PP    P+P
Sbjct:   639 PPPPTPVSSPPPPEKSPPPPPPAKSTPPPEEYPTPPTSVKSSPPPEKSLPPPTLIPSPPP 698

Query:    64 SAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTPALGPS 123
                        P K P      K      PP+E V +P  TP               V +P     PTP    P
Sbjct:   699 QEKPTPPSTPSKPPSSPEKPS-PPKEPVSSPPQTPK--SSPPPAPVSSP--PPTPVSSPP 753

Query:   124 APQKPLRRALSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSP 183
             A   P+     S ++SP PAP S     A ++K+    +    +  +   P PP    +    PP +P
Sbjct:   754 A-LAPVSSPPSVKSSPPPAPLSSPPPAPQVKS----SPPPVQVSSP--PPAPKSSPPLAP 806

Query:   184 ASTASFIFSKGSRKLQLERP-VSPETQADLQRNLVAELRSISEQRPPQAPKKSPKAPPPV 242
              S+   +      L  P ++P+++          +V+     + +   PP AP  SP  P
Sbjct:   807 VSSPPQVEKTSPPPAPLSSPPLAPKSSPP--HVVVSSPPPVVKSSPPPAPVSSPPLTPKP 864

Query:   243 ARKPS-VGVPP----PASPSYPR------AEPLTAPP 268
             A  P+ V  PP    P++P  P            +EP  ++PP
Sbjct:   865 ASPPAHVSSPPEVVKPSTPPAPTTVISPPSEPKSSPP 901

 Score = 189 (28.4 bits), Expect = 1.0e-11, P = 1.0e-11
 Identities = 86/271 (31%), Positives = 112/271 (41%)

Query:     5 PPPEEAFFSVASPEPAGPSGSPEL-VSSP--AASSSSATALQIQPPG--SPDPPPAP--- 56
             PPP   A  S    P  P    S  P + VSSP  A  SS    A    PP      PPPAP
Sbjct:   768 PPP--APLSSPPPAPQVKSSPPPVQVSSPPPAPKSSPPLAPVSSPPQVEKTSPPPAPLSS 825

Query:    57 PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAP 116
             P  AP SS P  V    P    PV   S       PP    V +P  +TP              V +P
Sbjct:   826 PPLAPKSSPPHVVVSSPP--PVVKSS---PPPAPVSSPPLTPKPASPPA--HVSSPPEVV 878

Query:   117 TPALGPSAPQKPLRRALSGRASPVPAPSSGLHAAVRLKAC-SLAASEGL---SSAQP--- 169
              P+  P AP    +      ++SP P P S        V+      ++ +S    +      SS  P
```

```
Sbjct:   879 KPST-PPAPTTVISPPSEPKSSPPPTPVSLPPPIVKSSPPPAMVSSPPMTPKSSPPPVVV 937

Query:   170 -NGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRP 228
              + PP  +  PP +P S+              + P  PE ++      V+      +  P
Sbjct:   938 SSPPPTVKSSPPPAPVSSPPATPKSSPPPAPVNLP-PPEVKSSPPPTPVSSPPPAPKSSP 996

Query:   229 PQAPKKSPKAPPPVARKPS----VGVPPPASPSYPRAEPLTAPP 268
             P AP  SP  PPP  + P     V  PPP   S P   P+++PP
Sbjct:   997 PPAPMSSP--PPPEVKSPPPPAPVSSPPPPVKSPPPPAPVSSPP 1038

 Score = 181 (27.2 bits), Expect = 8.8e-11, P = 8.8e-11
 Identities = 73/277 (26%), Positives = 105/277 (37%)

Query:     3 DFPPPEEAFFSVASPEPAGPSGSPELVSSPAASSSSATALQIQPP----GSPDPP---PA 55
             D+ PP      V    P   S SP+ V  PAAS+   + +++ PP    GSP PP    +
Sbjct:   469 DYVPPTPP---VPGKSPPATSPSPQ-VQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTS 524

Query:    56 PPAPAPASSAPGHVAKL----PQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGA 111
             PPAP  + S P  V+ +    P K P   +  G PP  +   P   P       ++S
Sbjct:   525 PPAPIGSPSPPPPVSVVSPPPPVKSPPPPAPVGSPPPPEKSPPPPAPVASPPPPVKSPPP 584

Query:   112 PG--GAPTPALGPSAPQKPLRRA---LSGRASPVPAPSSGLHAAVRLKACSLAASEGLSS 166
             P    +P P +    P  P+        +   P P  S    A V      + +    +
Sbjct:   585 PTLVASPPPPVKSPPPPAPVASPPPPVKSPPPPTPVASPPPPAPVASSPPPMKSPPPPTP 644

Query:   167 AQPNGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQ 226
                  PPE  P PP  PA +        +    ++    PE        L+       +
Sbjct:   645 VSSPPPPEKSP-PPPPPAKSTPPPEEYPTPPTSVKSSPPPEKSLP-PPTLIPSPPPQEKP 702

Query:   227 RPPQAPKKSPKAPP-PVARKPSVGVPPPASPSYPRAEPLTAPP 268
              PP  P K P +P  P   K  V  PP    S P   P+++PP
Sbjct:   703 TPPSTPSKPPSSPEKPSPPKEPVSSPPQTPKSSPPPAPVSSPP 745

 Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10
 Identities = 78/264 (29%), Positives = 105/264 (39%)

Query:     5 PPPEEAFFSVASPEPAGP----SGSPELVSSPAASSSSATALQIQPPGSP--DPPPAP-- 56
             PPP        +P+PA P    S  PE+V  P+   +  T   I PP  P   PPP P
Sbjct:   850 PPPAPVSSPPLTPKPASPPAHVSSPPEVVK-PSTPPAPTTV--ISPPSEPKSSPPPTPVS 906

Query:    57 -PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGA 115
              P P   SS P  +   P   P        PP   V +P   P++        V +P
Sbjct:   907 LPPPIVKSSPPPAMVSSPPMTPKS-----SPPPVVVSSP--PPTVKSSPPPAPVSSPPAT 959

Query:   116 PTPALGPSAPQKPLRRALSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEA 175
             P  +  P+    P       ++SP P P S    A +    S      +SS  P  PPE
Sbjct:   960 PKSSPPPAPVNLPPPEV---KSSPPPTPVSSPPPAPK----SSPPPAPMSSP-P--PPEV 1009

Query:   176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKS 235
             +  PP +P S+             +  P  P  ++      V+      +  PP AP  S
Sbjct:  1010 KSPPPPAPVSSPPPPVKSPPPPAPVSSP-PPPVKSPPPPAPVSSPPPPVKSPPPPAPISS 1068

Query:   236 PKAPPPVARKPS---VGVPPPASPSYPRAEPLTAPP 268
             P  PPPV   P    V  PPP   S P   P+++PP
Sbjct:  1069 P--PPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPP 1102

 Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10
 Identities = 82/267 (30%), Positives = 110/267 (41%)

Query:    17 PEPAG-PSGSPELVSSPAASS---SSATALQIQPPGSPDPPPAP---PAPAPASSAPGHV 69
             P P G P  SP +   PAAS+   S  T   + P   P+P P P   P P P   +P
Sbjct:   410 PTPGGGPPSSP-VPGKPAASAPMPSPHTPPDVSPEPLPEPSPVPAPAPMPMPTPHSPPAD 468

Query:    70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTPALGPSAPQKP 128
               +P   PV G S     P   V  P  +     +V+L    AP G+P P +  ++P  P
Sbjct:   469 DYVPPTPPVPGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTSPPAP 528

Query:   129 LRRALSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSPASTAS 188
             +     G  SP P P S +     +K+    A  G   + P  PPE  P PP   AS
Sbjct:   529 I-----GSPSP-PPPVSVVSPPPPVKSPPPPAPVG---SPP--PPEKSPPPPAPVASPPP 577

Query:   189 FIFSKGSRKLQLERPV---SPETQADLQRNLVAELRS------ISEQRPPQA------PK 233
              + S     L    P    SP   A +  +    ++S       ++  PP        P
Sbjct:   578 PVKSPPPPTLVASPPPPVKSPPPPAPVA-SPPPPVKSPPPPTPVASPPPPAPVASSPPPM 636

Query:   234 KSPKAPPPVARKP---SVGVPPPASPSYPRAEPLTAPPTN 270
             KSP  P PV+  P       PPP + S P  E    PPT+
Sbjct:   637 KSPPPPTPVSSPPPPEKSPPPPPPAKSTPPPEEYPTPPTS 676

 Score = 170 (25.5 bits), Expect = 1.6e-09, P = 1.6e-09
 Identities = 78/279 (27%), Positives = 108/279 (38%)
```

```
Query:     5 PPPEEAFFSVASPEPAGPSGSPELVSSPAASSSSATALQIQPPGSPDPPPAPPAPAPASS 64
             PP       S  S    + P  +P  +   P    SS    A+   PP +P     +PP P   SS
Sbjct:   883 PPAPTTVISPPSEPKSSPPPTPVSLPPPIVKSSPPPAMVSSPPMTPKS--SPP-PVVVSS 939

Query:    65 APGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPG--GAPTPALGP 122
             P  V   P   PV          PP      +P   P  L     ++S   P     +P PA
Sbjct:   940 PPPTVKSSPPPAPVS-----SPPATPKSSPPPAPVNLPPPEVKSSPPPTPVSSPPPAPKS 994

Query:   123 SAPQKPLRRALSG--RASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPP 180
             S P  P+        ++ P PAP S      V+      S         +SS  P  PP  +  PP
Sbjct:   995 SPPPAPMSSPPPPEVKSPPPPAPVSSPPPPVK----SPPPPAPVSS--P--PPPVKSPPP 1046

Query:   181 QSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKSPKAPP 240
             +P S+               +   P   P   ++        V+         +  PP AP   SP  PP
Sbjct:  1047 PAPVSSPPPPVKSPPPPAPISSP-PPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSP--PP 1103

Query:   241 PVARKPS---VGVPPPAS---PSYPRAEPLTAPPTNGLPHTQDRTKREL 283
             P+      P     V  PPPA         PS P     P+++PP       P        + ++ L
Sbjct:  1104 PIKSPPPPAPVSSPPPAPVKPPSLPPPAPVSSPPPVVTPAPPKKEEQSL 1152
```

Score = 169 (25.4 bits), Expect = 2.1e-09, P = 2.1e-09
Identities = 75/266 (28%), Positives = 104/266 (39%)

```
Query:     3 DFPPPEEAFFSVASPEPAGPSGSPELVSSPAASSSSATALQIQPP----GSPDPP---PA 55
             D+ PP         V       P      S SP+ V    PAAS+     + +++ PP        GSP PP         +
Sbjct:   469 DYVPPTPP---VPGKSPPATSPSPQ-VQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTS 524

Query:    56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGA 115
             PPAP    + S P    V+ +         PV              PP     VG+P      P                V +P
Sbjct:   525 PPAPIGSPSPPPPVSVVSPPPPVKSP----PPPAPVGSP--PPPEKSPPPPAPVASP--- 575

Query:   116 PTPALGPSAPQKPLRRALSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEA 175
             P P      P    P                        ++ P PAP +          V+          S                ++S    P      P    +
Sbjct:   576 PPPVKSPPPPTLVASPPPPVKSPPPPAPVASPPPPVK----SPPPPTPVASPPPPAPVAS 631

Query:   176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKS 235
             P  P  +SP                     K           P     P                                  S+          PP+
Sbjct:   632 SPPPMKSPPPPTPVSSPPPPEKSP--PPPPAKSTPPPEEYPTPPTSVKSSPPPEKSLPP 689

Query:   236 PK---APPPVARK--PSVGVPPPASPSYPRA--EPLTAPP 268
             P        +PPP      +      PS            PP+SP      P            EP+++PP
Sbjct:   690 PTLIPSPPPQEKPTPPSTPSKPPSSPEKPSPPKEPVSSPP 729
```

Score = 168 (25.2 bits), Expect = 2.7e-09, P = 2.7e-09
Identities = 75/267 (28%), Positives = 102/267 (38%)

```
Query:     2 ADFPPPEEAFFSVASPE-PAGPSGSPELVSSPAASSSSATALQIQPPGSPDPP-PAPPAP 59
             A   PPP   +    ++ P+ P G      P      +SP A      S +              SP PP   +PP P
Sbjct:   496 ASTPPP--SLVKLSPPQAPVGSPPPPVKTTSPPAPIGSPSPPPPVSVVSPPPPVKSPPPP 553

Query:    60 APASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTPA 119
             AP   S P          P     PV             PP +       P  +    S        V+          AP    +P  P
Sbjct:   554 APVGSPPPPEKSPPPPAPVASPP---PPVKSPPPPTLVASPPPPVKSPPPPAPVASPPPP 610

Query:   120 LGPSAPQKPLRRALSGRASPVPAPSSGLHAAVRLKACSL-AASEGLSSAQPNGPPEAEPR 178
             +       P   P+              + P PAP +          ++                +S                P   PP  A+
Sbjct:   611 VKSPPPPTPVA------SPPPPAPVASSPPPMKSPPPPTPVSSPPPPEKSPPPPPPAKST 664

Query:   179 PP--QSPASTASFIFSKGSRKLQLERPV---SPETQADLQRNLVAELRSISEQRPPQAPK 233
             PP   + P       S      S          K    L    P           SP   Q                           S  ++P     +P
Sbjct:   665 PPPEEYPTPPTSVKSSPPPEK-SLPPPTLIPSPPPQEKPTPPSTPSKPPSSPEKP--SPP 721

Query:   234 KSPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
             K  P  + PP      K  S          PPPA    S  P        P+++PP
Sbjct:   722 KEPVSSPPQTPKSS---PPPAPVSSPPPTPVSSPP 753
```

Score = 166 (24.9 bits), Expect = 4.6e-09, P = 4.6e-09
Identities = 81/268 (30%), Positives = 108/268 (40%)

```
Query:     5 PPPEEAF---FSVASPEPAGPSGSPE-LVSSPAASSSS----ATALQIQPPGSPDPPP-- 54
             PPPE++             VASP  P     S   P   LV+SP       S         A              PP       PPP
Sbjct:   560 PPPEKSPPPPAPVASPPPPVKSPPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPPTP 619

Query:    55 --APPAPAPASSAPGHVAKLPQKEPVGC----SKGGGPPREDVGAPLVTPSLLQMVRLRS 108
               +PP  PAP   +S+P     +       P       PV                  K       PP                 P                                ++S
Sbjct:   620 VASPPPPAPVASSPPPMKSPPPPTPVSSPPPPEKSPPPPPPAKSTPPPEEYPTPPTSVKS 679

Query:   109 VGAPGGA-PTPALGPSAPQKPLRRALSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSA 167
                 P     +    P     P   L   PS    P      P         +         +     ++P            PSS                +         +     S                            SS
Sbjct:   680 SPPPEKSLPPPTLIPSPP--PQEKP-TPPSTPSKPPSSPEKPSPPKEPVSSPPQTPKSSP 736
```

```
Query:   168 QPNGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQR 227
              P       P P  SP + A  + S  S K     P +P +       + +    +
Sbjct:   737 PPAPVSSPPPTPVSSPPALAP-VSSPPSVKSS--PPPAPLSSPPPAPQVKSSPPPVQVSS 793

Query:   228 PPQAPKKSPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
             PP APK SP    P+A  P V  PP    + P   PL++PP
Sbjct:   794 PPPAPKSSP----PLA--P-VSSPPQVEKTSPPPAPLSSPP 827

 Score = 165 (24.8 bits), Expect = 6.0e-09, P = 6.0e-09
 Identities = 79/264 (29%), Positives = 105/264 (39%)

Query:     5 PPPEEAFFSVASPEPAG-PSGSP--ELVSSPAASSSSATALQIQPPGSPDPPP-APPAPA 60
             PPP +   + + P P G PS  P   +VS P    S        P GSP PP  +PP PA
Sbjct:   517 PPPVK---TTSPPAPIGSPSPPPPVSVVSPPPPVKSPPPPA---PVGSPPPPEKSPPPPA 570

Query:    61 PASSAPGHVAKLPQKEPVGCSKG---GGPPREDVGAP---LVTPSLLQMVRLRSVGAPGG 114
             P +S P  V   P    V          PP   V +P   + +P     V      AP
Sbjct:   571 PVASPPPPVKSPPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPPTPVASPPPPAPVA 630

Query:   115 APTPALGPSAPQKPLRRALSGRASPVPAP---SSGLHAAVRLKACSLAASEGLSSAQPNG 171
             +  P +    P  P+        SP P P   S+           S+ +S     + P
Sbjct:   631 SSPPPMKSPPPPTPVSSPPPPEKSPPPPPPAKSTPPPEEYPTPPTSVKSSPPPEKSLP-- 688

Query:   172 PPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQA 231
             PP   P PP     T     SK         P SPE  +   +  V+      +  PP A
Sbjct:   689 PPTLIPSPPPQEKPTPPSTPSKP--------PSSPEKPSP-PKEPVSSPPQTPKSSPPPA 739

Query:   232 PKKSPKAPPPVARKPSVGV--PPPASPSYPRAEPLTAPP 268
             P  SP  P PV+  P++     PP+  S P   PL++PP
Sbjct:   740 PVSSPP-PTPVSSPPALAPVSSPPSVKSSPPPAPLSSPP 777

 Score = 162 (24.3 bits), Expect = 1.3e-08, P = 1.3e-08
 Identities = 76/272 (27%), Positives = 99/272 (36%)

Query:     2 ADFPPPEEAFFSVASPEPAG-PSGSPELVSSPAASSSSATALQIQPPGSPDPPPAPPAPA 60
             A  P P        SPEP   PS  P     P  +  S  A    PP  P P  +PPA +
Sbjct:   427 ASAPMPSPHTPPDVSPEPLPEPSPVPAPAPMPMPTPHSPPADDYVPPTPPVPGKSPPATS 486

Query:    61 PASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTP-- 118
             P+       A  P    V  S    PP+  VG+P   P     V+  S  AP G+P+P
Sbjct:   487 PSPQVQPPAASTPPPSLVKLS----PPQAPVGSP--PPP----VKTTSPPAPIGSPSPPP 536

Query:   119 ---ALGPSAPQK-PLRRALSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174
                 + P  P K P   A  G  SP P   S    A          S    +   + PP
Sbjct:   537 PVSVVSPPPPVKSPPPPAPVG--SPPPPEKSPPPPAPVASPPPPVKSPPPPTLVASPPPP 594

Query:   175 AEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKK 234
              +  PP +P ++             +  P  P   A     + +         PP  P+K
Sbjct:   595 VKSPPPPAPVASPPPPVKSPPPPTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPP-PEK 653

Query:   235 SPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPPTNGLP 273
             SP  PPP    P    PP   P+ P +   + PP   LP
Sbjct:   654 SPPPPPPAKSTP----PPEEYPTPPTSVKSSPPPEKSLP 688

 Score = 159 (23.9 bits), Expect = 2.8e-08, P = 2.8e-08
 Identities = 77/264 (29%), Positives = 103/264 (39%)

Query:     5 PPPEEAFFSVASPEPAGPSGSPELVSSPAASSSSATALQIQPPGSP--DPPPAP---PAP 59
             PPP      V+SP P  P  SP     P   SS    ++  PP +P   PP  P   P P
Sbjct:   916 PPPA----MVSSP-PMTPKSSPP----PVVVSSPPPTVKSSPPPAPVSSPPATPKSSPPP 966

Query:    60 APASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTPA 119
             AP +  P  V   P   PV  S    P      AP+ +P   + V+     AP  +P P
Sbjct:   967 APVNLPPPEVKSSPPPTPVS-SPPPAPKSSPPPAPMSSPPPPE-VKSPPPPAPVSSPPPP 1024

Query:   120 LGPSAPQKPLRRALSG-RASPVPAPSSGLHAAVRLKACSLAASEG---LSSAQPNGPPEA 175
             +    P  P+       ++ P PAP S     V+        S     + S  P  P  +
Sbjct:  1025 VKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPISSPPPPVKSPPPPAPVSS 1084

Query:   176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKS 235
              P P +SP   A       S    ++ P  P   A +     A ++  S   PP AP  S
Sbjct:  1085 PPPPVKSPPPPAPV----SSPPPPIKSPPPP---APVSSPPPAPVKPPS--LPPPAPVSS 1135

Query:   236 PK--APPPVARKPSVGVPPPA-SPSYPRAEPLTAPP 268
             P     P   +K    +PPPA S   P    +  PP
Sbjct:  1136 PPPVVTPAPPKKEEQSLPPPAESQPPPSFNDIILPP 1171

 Score = 143 (21.5 bits), Expect = 1.8e-06, P = 1.8e-06
 Identities = 59/179 (32%), Positives = 77/179 (43%)
```

```
Query:     3 DFPPPEEAFFSVASPEP-AGPSGSPELVSSPAASSSSATA-LQIQPPGSP--DPPP---A 55
             + PPPE     S    P P + P  +P+     PA  SS     ++  PP +P    PPP   +
Sbjct:   970 NLPPPEVK--SSPPPTPVSSPPPAPKSSPPPAPMSSPPPPEVKSPPPPAPVSSPPPPVKS 1027

Query:    56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGA 115
             PP PAP SS P  V    P    PV          PP +      P      S      V+      AP  +
Sbjct:  1028 PPPPAPVSSPPPPVKSPPPPAPVSSPP---PPVKSPPPPAPISSPPPPVKSPPPPAPVSS 1084

Query:   116 PTPALGPSAPQKPLRRALSG-RASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174
             P P +     P  P+        ++ P PAP S      A  +K  SL       +SS  P  PP
Sbjct:  1085 PPPPVKSPPPPAPVSSPPPPIKSPPPPAPVSSPPPAP-VKPPSLPPPAPVSS--P--PPV 1139

Query:   175 AEPRPPQ 181
                P PP+
Sbjct:  1140 VTPAPPK 1146

 Score = 133 (20.0 bits), Expect = 2.3e-05, P = 2.3e-05
 Identities = 50/132 (37%), Positives = 59/132 (44%)

Query:     1 MADFPPPEEAFFSVASPEPAGP-SGSPELVSSP---AASSSSATALQIQPPGSP--DPPP 54
             M+  PPPE      V SP P  S  P  V SP    A  SS      ++  PP +P    PPP
Sbjct:  1001 MSSPPPPE-----VKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPP 1055

Query:    55 ---APPAPAPASSAPGHVAKLPQKEPVGCSKG---GGPPREDVGAPLVTPSLLQMVRLRS 108
                +PP PAP SS P   V    P    PV              PP      V +P    P +
Sbjct:  1056 PVKSPPPPAPISSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSP--PPPIKSPPPPAP 1113

Query:   109 VGAPGGAPT--PALGPSAP 125
             V +P   AP     P+L P AP
Sbjct:  1114 VSSPPPAPVKPPSLPPPAP 1132

 Score = 110 (16.5 bits), Expect = 8.0e-03, P = 8.0e-03
 Identities = 41/121 (33%), Positives = 49/121 (40%)

Query:     5 PPPEEAFFS----VASPEPAGP-SGSPELVSSP---AASSSSATALQIQPPGSP--DPPP 54
             PPP      S      V SP P   P S   P V SP    A   SS      ++  PP +P    PPP
Sbjct:  1060 PPPPAPISSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPIKSPPPPAPVSSPPP 1119

Query:    55 AP------PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRS 108
             AP        P PAP SS P   V      P K+         +     PP E        P         +L    +
Sbjct:  1120 APVKPPSLPPPAPVSSPPPVVTPAPPKKE---EQSLPPPAESQPPPSFNDIILPPIMANK 1176

Query:   109 VGAP 112
               +P
Sbjct:  1177 YASP 1180

 Score = 108 (16.2 bits), Expect = 1.3e-02, P = 1.3e-02
 Identities = 46/155 (29%), Positives = 67/155 (43%)

Query:   114 GAPTPALGPSAPQKPLRRALSGRASPVPAPSSGLHAAVR-LKACS-LAASEGLSSAQPNG 171
             G PTP  GP +    P + A S   +P+P+P +      +     L      S  + A      +     P+
Sbjct:   408 GYPTPGGGPPSSPVPGKPAAS---APMPSPHTPPDVSPEPLPEPSPVPAPAPMPMPTPHS 464

Query:   172 PPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQ----ADLQRNLVAELRSISEQR 227
             PP   +  PP  P        S     +  S ++Q           +P           +  Q   + +    +
Sbjct:   465 PPADDYVPPTPPVPGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTS 524

Query:   228 PPQAPKKSPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268
             PP AP   SP   PPPV        SV   PPP      S P    P+  +PP
Sbjct:   525 PP-APIGSPSPPPPV----SVVSPPPPVKSPPPPAPVGSPP 560
```

Pedant information for DKFZphmcf1_1c23, frame 1

Report for DKFZphmcf1_1c23.1

```
[LENGTH]    311
[MW]        31534.58
[pI]        9.48
[KW]        All_Alpha
[KW]        LOW_COMPLEXITY    38.59 %

SEQ    MADFPPPEEAFFSVASPEPAGPSGSPELVSSPAASSSSATALQIQPPGSPDPPPAPPAPA
SEG    ...........xxxxxxxxxxxxxxx..xxxxxxxxxxxx....xxxxxxxxxxxxxxx
PRD    cccccccccccccccccccccccccccccccccccccccceeeeccccccccccccccccc

SEQ    PASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPTPAL
SEG    xxxxxx..........................................xxxxxxxxxxx
```

```
PRD   ccccccccccccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccc

SEQ   GPSAPQKPLRRALSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPP
SEG   xxxxx..........................................xxxxxxxxxxxxx
PRD   cccccchhhhhhhhcccccccccccchhhhhhhhhhhhhhhccccccccccccccccccc

SEQ   QSPASTASFIFSKGSRKLQLERPVSPETQADLQRNLVAELRSISEQRPPQAPKKSPKAPP
SEG   xxxxx........................................xxxxxxxxxxxxxxx
PRD   ccccccceeeecccccchhhhcccccccchhhhhhhhhhhhhhhhccccccccccccccccc

SEQ   PVARKPSVGVPPPASPSYPRAEPLTAPPTNGLPHTQDRTKRELAENGGVLQLVGPEEKMG
SEG   xxxxxxxxxxxxxxxxxxxxxxx.....................................
PRD   ccccccccccccccccccccccccccccccccccccchhhhhhcccceeecccccccccc

SEQ   LPGSDSQKELA
SEG   ...........
PRD   ccccccccccc
```

(No Prosite data available for DKFZphmcf1_1c23.1)

(No Pfam data available for DKFZphmcf1_1c23.1)

DKFZphmcf1_1e15
----

group: transmembrane protein

DKFZphmcf1_1e15 encodes a novel 454 amino acid protein with similarity to C. elegans proteins and transporter proteins.

The novel protein is similar to the PTR2 family of proton/oligopeptide symporter proteins and the D-xylose-proton symporter. Thus, the protein is a transporter of a so far unknown compound.

The new protein can find application as a new transporter in eukaryotic cells, e.g. in drug transport into cells.

similarity to D-XYLOSE TRANSPORTER
membrane regions: 9

complete cDNA, complete cds, EST hits
matchs cDNA encoding cell growth inhibiting factor (E12646)

Sequenced by DKFZ

Locus: unknown

Insert length: 1957 bp
Poly A stretch at pos. 1947, polyadenylation signal at pos. 1929

```
   1 GGTGCAGCGC CCGGGCTGAG CGACAGCAAG TGCAGCGGGC TCCTACCCCG
  51 GGTGAGGGGT GGCCTCCGCG TGGGATCGTG CCCTCTTCAG CCCGCTCCTG
 101 TCCCCGACAT CACGTGTATT CCGCACGTCC CCTCCGCGCT GTGTGTCTAC
 151 TGAGACGGGG AGGCGTGACA GGGCCCGGGT CCCTTCTCAG TGGTGCTCTG
 201 TGCTTCAGGG CAAGCTCCCC GTCTCCGGGC GCACTTCCCT CGCCTGTGTT
 251 CGGTCCATCC TCCTTTCTCC AGCCTCCTCC CCTCGCAGGT GGGATCGTCG
 301 GTGGGACCGG AGCGCGGGCG GGCGCGGCCC CCCGGGACCA TGGCCGGGTC
 351 CGACACCGCG CCCTTCCTCA GCCAGGCGGA TGACCCGGAC GACGGGCCAG
 401 TGCCTGGCAC CCCGGGGTTG CCAGGGTCCA CGGGGAACCC GAAGTCCGAG
 451 GAGCCCGAGG TCCCGGACCA GGAGGGGCTG CAGCGCATCA CCGGCCTGTC
 501 TCCCGGCCGT TCGGCTCTCA TAGTGGCGGT GCTGTGCTAC ATCAATCTCC
 551 TGAACTACAT GGACCGCTTC ACCGTGGCTG TGTTCATCTC CAGTTACATG
 601 GTGTTGGCAC CTGTGTTTGG CTACCTGGGT GACAGGTACA ATCGGAAGTA
 651 TCTCATGTGC GGGGGCATTG CCTTCTGGTC CCTGGTGACA CTGGGGTCAT
 701 CCTTCATCCC CGGAGAGCAT TTCTGGCTGC TCCTCCTGAC CCGGGGCCTG
 751 GTGGGGGTCG GGGAGGCCAG TTATTCCACC ATCGCGCCCA CTCTCATTGC
 801 CGACCTCTTT GTGGCCGACC AGCGGAGCCG GATGCTCAGC ATCTTCTACT
 851 TTGCCATTCC GGTGGGCAGT GGTCTGGGCT ACATTGCAGG CTCCAAAGTG
 901 AAGGATATGG CTGGAGACTG GCACTGGGCT CTGAGGGTGA CACCGGGTCT
 951 AGGAGTGGTG GCCGTTCTGC TGCTGTTCCT GGTAGTGCGG GAGCCGCCAA
1001 GGGGAGCCGT GGAGCGCCAC TCAGATTTGC CACCCCTGAA CCCCACCTCG
1051 TGGTGGGCAG ATCTGAGGGC TCTGGCAAGA AATCTCATCT TTGGACTCAT
1101 CACCTGCCTG ACCGGAGTCC TGGGTGTGGG CCTGGGTGTG GAGATCAGCC
1151 GCCGGCTCCG CCACTCCAAC CCCCGGGCTG ATCCCCTGGT CTGTGCCACT
1201 GGCCTCCTGG GCTCTGCACC CTTCCTCTTC CTGTCCCTTG CCTGCGCCCG
1251 TGGTAGCATC GTGGCCACTT ATATTTTCAT CTTCATTGGA GAGACCCTCC
1301 TGTCCATGAA CTGGGCCATC GTGGCCGACA TTCTGCTGTA CGTGGTGATC
1351 CCTACCCGAC GCTCCACCGC CGAGGCCTTC CAGATCGTGC TGTCCCACCT
1401 GCTGGGTGAT GCTGGGAGCC CCTACCTCAT TGGCCTGATC TCTGACCGCC
1451 TGCGCCGGAA CTGGCCCCCC TCCTTCTTGT CCGAGTTCCG GGCTCTGCAG
1501 TTCTCGCTCA TGCTCTGCGC GTTTGTTGGG GCACTGGGCG GCGCAGCCTT
1551 CCTGGGCACC GCCATCTTCA TTGAGGCCGA CCGCCGGCGG GCACAGCTGC
1601 ACGTGCAGGG CCTGCTGCAC GAAGCAGGGT CCACAGACGA CCGGATTGTG
1651 GTGCCCCAGC GGGGCCGCTC CACCCGCGTG CCCGTGGCCA GTGTGCTCAT
1701 CTGAGAGGCT GCCGCTCACC TACCTGCACA TCTGCCACAG CTGGCCCTGG
1751 GCCCACCCCA CGAAGGGCCT GGGCCTAACC CCTTGGCCTG GCCCAGCTTC
1801 CAGAGGGACC CTGGGCCGTG TGCCAGCTCC CAGACACTAC ATGGGTAGCT
1851 CAGGGGAGGA GGTGGGGGTC CAGGAGGGGG ATCCCTCTCC ACAGGGGCAG
1901 CCCCAAGGGC TCGGTGCTAT TTGTAACGGA ATAAAATTTG TAGCCAGAAA
1951 AAAAAAA
```

BLAST Results

Entry E12646 from database EMBL:
cDNA encoding cell growth inhibiting factor.
Score = 3046, P = 2.2e-131, identities = 640/659

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 340 bp to 1701 bp; peptide length: 454
Category: similarity to known protein

```
  1 MAGSDTAPFL SQADDPDDGP VPGTPGLPGS TGNPKSEEPE VPDQEGLQRI
 51 TGLSPGRSAL IVAVLCYINL LNYMDRFTVA VFISSYMVLA PVFGYLGDRY
101 NRKYLMCGGI AFWSLVTLGS SFIPGEHFWL LLLTRGLVGV GEASYSTIAP
151 TLIADLFVAD QRSRMLSIFY FAIPVGSGLG YIAGSKVKDM AGDWHWALRV
201 TPGLGVVAVL LLFLVVREPP RGAVERHSDL PPLNPTSWWA DLRALARNLI
251 FGLITCLTGV LGVGLGVEIS RRLRHSNPRA DPLVCATGLL GSAPFLFLSL
301 ACARGSIVAT YIFIFIGETL LSMNWAIVAD ILLYVVIPTR RSTAEAFQIV
351 LSHLLGDAGS PYLIGLISDR LRRNWPPSFL SEFRALQFSL MLCAFVGALG
401 GAAFLGTAIF IEADRRRAQL HVQGLLHEAG STDDRIVVPQ RGRSTRVPVA
451 SVLI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1_1e15, frame 1

TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4, N = 3, Score = 441, P = 5.2e-76

TREMBL:CEC39E9_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid C39E9, N = 2, Score = 449, P = 8.2e-69

TREMBL:CEF09A5_1 gene: "F09A5.1"; Caenorhabditis elegans cosmid F09A5, N = 3, Score = 413, P = 9.1e-60

TREMBL:ATF6H11_18 gene: "F6H11.180"; product: "predicted protein"; Arabidopsis thaliana DNA chromosome 5, BAC clone F6H11 (ESSAII project), N = 3, Score = 193, P = 2.5e-24

SWISSPROT:XYLT_LACBR D-XYLOSE-PROTON SYMPORT (D-XYLOSE TRANSPORTER)., N = 1, Score = 180, P = 7.9e-11

```
>TREMBL:CEC39E9_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid C39E9
          Length = 488

  HSPs:

 Score = 449 (67.4 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69
 Identities = 88/204 (43%), Positives = 125/204 (61%)

Query:    58 SALIVAVLCYINLLNYMDRFTVAVFISSYMVLAPVFGYLGDRYNRKYLMCGGIAFWSLVT 117
             + ++  V  Y N+ + +       VF+ S+MV +PV GYLGDR+NRK++M  G+  W
Sbjct:    29 AGVLTQVQTYYNISDSLGGLIQTVFLISFMVFSPVCGYLGDRFNRKWIMIIGVGIWLGAV 88

Query:   118 LGSSFIPGEHFWLLLLTRGLVGVGEASYSTIAPTLIADLFVADQRSRMLSIFYFAIPVGS 177
             LGSSF+P  HFWL L+ R  VG+GEASYS +AP+LI+D+F    +RS +  IFYFAIPVGS
Sbjct:    89 LGSSFVPANHFWLFLVLRSFVGIGEASYSNVAPSLISDMFNGQKRSTVFMIFYFAIPVGS 148

Query:   178 GLGYIAGSKVKDMAGDWHWALRVTPGLGVVAVLLLFLVVREPPRGAVER----HSDLPPL 233
             GLG+I GS V   + G W W +RV+   G++ ++ L L   EP RGA ++        D+
Sbjct:   149 GLGFIVGSNVATLTGHWQWGIRVSAIAGLIVMIALVLFTYEPERGAADKAMGESKDVVVT 208

Query:   234 NPTSWWADLRALARNLIFGLITCLTG 259
              T++  DL  L +      L+ C  G
Sbjct:   209 TNTTYLEDLVILLKTPT--LVACTWG 232

 Score = 267 (40.1 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69
 Identities = 74/212 (34%), Positives = 113/212 (53%)

Query:   249 LIFGLITCLTGVLGVGLGVEISRRL--------RHSNPRADPLVCATGLLGSAPFLFLSL 300
             L FG IT   G++GV  G  +S+ L         R    RA PLV   G L +APFL + +
Sbjct:   277 LYFGAITTAGGLIGVIFGSMLSKWLVAGWGPFRRLQTDRAQPLVAGGGALLAAPFLLIGM 336

Query:   301 ACARGSIVATYIFIFIGETLLSMNWAIVADILLYVVIPTRRSTAEAFQIVLSHLLGDAGS 360
```

```
                 S+V  YI IF G T +  NW +  D+L  V+ P RRSTA ++ +++SHL GDA
Sbjct:   337 IFGDKSLVLLYIMIFFGITFMCFNWGLNIDMLTTVIHPNRRSTAFSYFVLVSHLFGDASG 396

Query:   361 PYLIGLISDRLRRN--WPPSFLSEFRALQFSLMLCAFVGALGGAAFLGTAIFIEADRR-- 416
             PYLIGLISD +R    +P     ++ +L  +    C  +  L     +  +++ + +DR+
Sbjct:   397 PYLIGLISDAIRHGSTYPKD---QYHSLVSATYCCVALLLLSAGLYFVSSLTLVSDRKKF 453

Query:   417 RAQLHVQGLLHEA--GSTD--DRIVVPQRGRSTRV 447
             RA++ +  L  +    STD  +RI +      S+R+
Sbjct:   454 RAEMGLDDLQSKPIRTSTDSLERIGINDDVASSRL 488

 Score = 70 (10.5 bits), Expect = 5.9e-24, Sum P(2) = 5.9e-24
 Identities = 25/89 (28%), Positives = 41/89 (46%)

Query:    62 VAVLCYINLLNYMDRFTVAVFISSYMVLAPVFGYLGDRYNRKYLMCGGIAFWSLVT--LG 119
             V  L  +NLLNY+DR+TVA  ++         +   LG       +L+      +S V   LG
Sbjct:    11 VTALFVVNLLNYVDRYTVAGVLTQVQTYYNISDSLGGLIQTVFLI--SFMVFSPVCGYLG 68

Query:   120 SSFIPGEHFWLLLLTRGLVGVGEASYSTIAP 150
               F      W++++  G + +G    S+  P
Sbjct:    69 DRF---NRKWIMIIGVG-IWLGAVLGSSFVP 95
```

Pedant information for DKFZphmcf1_1e15, frame 1

Report for DKFZphmcf1_1e15.1

```
[LENGTH]     454
[MW]         49013.35
[pI]         7.66
[HOMOL]      TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4 2e-51

[BLOCKS]     BL01022D
[PROSITE]    MYRISTYL            11
[PROSITE]    CAMP_PHOSPHO_SITE    1
[PROSITE]    CK2_PHOSPHO_SITE     3
[PROSITE]    PROKAR_LIPOPROTEIN   1
[PROSITE]    GLYCOSAMINOGLYCAN    1
[PROSITE]    PKC_PHOSPHO_SITE     4
[KW]         TRANSMEMBRANE  8
[KW]         LOW_COMPLEXITY     15.42 %

SEQ    MAGSDTAPFLSQADDPDDGPVPGTPGLPGSTGNPKSEEPEVPDQEGLQRITGLSPGRSAL
SEG    ..............xxxxxxxxxxxxxxx...............................
PRD    ccccccceeeeecccccccccccccccccccccccccccccccccccceeeecccccchhhh
MEM    ......................................MMMMMMMMMMMMMMMMMMMMMMM

SEQ    IVAVLCYINLLNYMDRFTVAVFISSYMVLAPVFGYLGDRYNRKYLMCGGIAFWSLVTLGS
SEG    ............................................................
PRD    hhhhhhhhccccccccceeeeeehhhheeeeccccccccccceeeeeeecccceeeeeecc
MEM    MMMMMM.............MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM............

SEQ    SFIPGEHFWLLLLTRGLVGVGEASYSTIAPTLIADLFVADQRSRMLSIFYFAIPVGSGLG
SEG    .........xxxxxxxxxxxx.......................................
PRD    cccccchhhhhhhhhhcccccceeeeecceeeccccccccchhhheeeeeeccccce
MEM    ...........MMMMMMMMMMMMMMMM.................MMMMMMMMMMMMMMMM

SEQ    YIAGSKVKDMAGDWHWALRVTPGLGVVAVLLLFLVVREPPRGAVERHSDLPPLNPTSWWA
SEG    ......................xxxxxxxxxxxx........................
PRD    eeeccccccccccccceeeeeeecchhhhhhhhhhhcccccchhhhhcccccccccchh
MEM    MMMMMMMM....................................................

SEQ    DLRALARNLIFGLITCLTGVLGVGLGVEISRRLRHSNPRADPLVCATGLLGSAPFLFLSL
SEG    ...........xxxxxxxxxxxxxxx................................
PRD    hhhhhhhhhhhhheeeecccceeehhhhhhhhhcccccccceeccccceeeecccceeec
MEM    ...................................MMMMMMMMMMMMMMMMMMMMMMMMM

SEQ    ACARGSIVATYIFIFIGETLLSMNWAIVADILLYVVIPTRRSTAEAFQIVLSHLLGDAGS
SEG    ............................................................
PRD    cccchhhhheeeeeecccccccccchhhhhhheeeeeccccchhhhhhccccccccccccc
MEM    MMMM..............MMMMMMMMMMMMMMMMMMMMM.......MMMMMMMMMMMMM

SEQ    PYLIGLISDRLRRNWPPSFLSEFRALQFSLMLCAFVGALGGAAFLGTAIFIEADRRRAQL
SEG    ...............................xxxxxxxxxxxxx..............
PRD    ceeehhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhcccccceeeeeehhhhhhhh
MEM    MMMMMMM.................................................MM

SEQ    HVQGLLHEAGSTDDRIVVPQRGRSTRVPVASVLI
```

```
SEG   .................................
PRD   hhhhhhhhccccceeeeeecccccccceeeeeccc
MEM   MMMMMMMMMMMMMMMMMMMMMMMMMMMM.......
```

Prosite for DKFZphmcf1_1e15.1

PS00002	177->181	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	340->344	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	270->273	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	368->371	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00006	11->15	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	431->435	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00008	252->258	MYRISTYL	PDOC00008
PS00008	262->268	MYRISTYL	PDOC00008
PS00008	266->272	MYRISTYL	PDOC00008
PS00008	288->294	MYRISTYL	PDOC00008
PS00008	305->311	MYRISTYL	PDOC00008
PS00008	397->403	MYRISTYL	PDOC00008
PS00013	292->303	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphmcf1_1e15.1)

DKFZphmcf1_1g13
----

group: mammary carcinoma derived

DKFZphmcf1_1g13 encodes a novel 573 amino acid protein with very weak similarity to the human KIAA0543 protein and Musca domestica hermes transposase.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes.

similarity to KIAA0766

commplete cDNA, complete cds, few EST hits
on genomic level encoded by AC005020, no splicing, genomic?

Sequenced by DKFZ

Locus: unknown

Insert length: 2210 bp
Poly A stretch at pos. 2200, polyadenylation signal at pos. 2176

```
   1 GAAACCTGAT CTCATAAAAC CTAGGTCACA AAGGACAGCC CTGCAAAACA
  51 GACCCTATTT GGATCAAGTG AGCCAGTTCC TGGAACCTGA ATAATGACTC
 101 CTGAATCAAG GGATACTACA GATTTGTCTC CAGGGGGTAC CCAGGAGATG
 151 GAAGGCATCG TGATAGTGAA GGTGGAGGAG GAAGATGAAG AAGACCATTT
 201 TCAAAAGGAA AGAAACAAAG TAGAGTCATC GCCACAAGTT CTCAGTCGCT
 251 CTACAACTAT GAATGAGAGA GCCTTATTGT CATCGTATTT AGTTGCATAT
 301 AGAGTGGCAA AAGAGAAAAT GGCTCACACA GCGGCTGAAA AAATTATCCT
 351 TCCAGCATGT ATGGACATGG TACGGACAAT TTTTGATGAC AAATCAGCTG
 401 ATAAACTAAG AACTATACCT CTTAGTGATA ATACAATATC TCGTCGAATC
 451 TGTACGATTG CAAAACATTT GGAAGCAATG CTTATTACAC GGCTGCAGTC
 501 CGGTATAGAC TTTGCAATCC AACTCGATGA GAGCACTGAT ATTGCAAGTT
 551 GTCCCACACT CTTGGTTTAT GTCAGATATG TGTGGCAAGA TGATTTTGTA
 601 GAGGATCTCT TATGTTGTTT AAATTTAAAT TCACATATAA CTGGATTAGA
 651 TTTATTTACT GAATTAGAAA ACTGCCTTCT TGGTCAGTAT AAATTAAACT
 701 GGAAACATTG TAAAGGAATT TCAAGTGATG GAACAGCAAA TATGACCGGA
 751 AAACACAGCA GACTTACTGA AAAATTGTTA GAAGCAACCC ACAACAATGC
 801 TGTTTGGAAT CACTGTTTTA TTCATCGAGA AGCTTTGGTA TCCAAAGAAA
 851 TTTCACCAAG TCTGATGGAT GTATTGAAAA ATGCAGTGAA AACTGTTAAT
 901 TTTATTAAAG GAAGCTCACT GAATAGCCGA CTTCTCGAAA TATTTTGTTC
 951 AGAGATTGGA GTGAACCACA CCCACTTATT GTTTCATACA GAAGTTCGTT
1001 GGCTTTCTCA AGGAAAAGTA TTGAGCAGAG TATATGAACT CAGGAACGAG
1051 ATTTACATTT TTCTCGTTGA AAAGCAATCT CATTTGGCAA ATATTTTTGA
1101 AGACGACATT TGGGTAACAA AATTGGCATA TTTAAGTGAT ATTTTTGGCA
1151 TTCTTAATGA ATTAAGCCTG AAAATGCAGG GGAAAAACAA TGATATATTT
1201 CAGTATCTTG AACATATTCT AGGATTCCAA AAGACGTTAT TATTGTGGCA
1251 AGCAAGACTT AAAAGTAACC GCCCTAGCTA CTATATGTTT CCAACATTAT
1301 TGCAACACAT CGAAGAGAAC ATTATTAATG AAGACTGCTT AAAAGAAATA
1351 AAATTAGAGA TATTGTTGCA TCTCACTTCT TTGTCTCAAA CTTTTAATTA
1401 TTACTTTCCG GAAGAGAAAT TTGAATCATT AAAGGAAAAT ATTTGGATGA
1451 AAGATCCATT TGCTTTTCAA AACCCAGAAT CAATAATTGA GTTAAACTTG
1501 GAGCCTGAAG AAGAGAATGA ATTATTGCAG CTCAGTTCAT CATTCACACT
1551 AAAGAATTAT TATAAGATAT TAAGTTTATC AGCATTTTGG ATTAAGATTA
1601 AAGATGACTT TCCACTGCTA AGTAGGAAGA GTATATTGCT GTTACTACCA
1651 TTCACAACTA CATATTTGTG TGAACTAGGA TTTTCAATCT TGACACGGTT
1701 AAAAACAAAG AAGAGAAATA GGCTCAATAG TGCACCAGAT ATGCGGGTAG
1751 CATTATCTTC ATGTGTTCCT GACTGGAAGG AACTTATGAA CAGACAAGCA
1801 CACCCATCAC ATTAAATACA AACTTTACAA AATTCTGTGT ATAGCCAGGT
1851 GTGGTGGCTT ACGCCTGTAA TCCCAGCAGT GGGAGACCGA GGTGGGCAGA
1901 TCACTTGAGT TCAAGACCAG CCTGGCCAAC ATGGTGAAAC CCCATCTCTA
1951 CTAAAAATAG AAACCTTAGC CAGGCGTGGT GGCACATGCC TGCAGTCCCA
2001 GTTACTTGGG TGCCTGAGGC AGGAGAATCT CTTAAACCAG GAAGGCAGAG
2051 ATTGCAGTGA GCTGAGATAA TCCCACTGCA TTCCAGCCTG GGCAACAGCG
2101 TGAGACTTCA TCTCAAAAAA AAAAAATTGT ATTTGTACTT TTAAAGGGAT
2151 TTTGCAGTAT GTTGTAGTTA AACGTTAATA AAATTATATT TGTAATTAGG
2201 AAAAAAAAAA
```

BLAST Results

Entry AC005020 from database EMBL:
Homo sapiens clone GS259H13; HTGS phase 1, 4 unordered pieces.
Score = 9110, P = 0.0e+00, identities = 1822/1822

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 94 bp to 1812 bp; peptide length: 573
Category: similarity to unknown protein

```
  1 MTPESRDTTD LSPGGTQEME GIVIVKVEEE DEEDHFQKER NKVESSPQVL
 51 SRSTTMNERA LLSSYLVAYR VAKEKMAHTA AEKIILPACM DMVRTIFDDK
101 SADKLRTIPL SDNTISRRIC TIAKHLEAML ITRLQSGIDF AIQLDESTDI
151 ASCPTLLVYV RYVWQDDFVE DLLCCLNLNS HITGLDLFTE LENCLLGQYK
201 LNWKHCKGIS SDGTANMTGK HSRLTEKLLE ATHNNAVWNH CFIHREALVS
251 KEISPSLMDV LKNAVKTVNF IKGSSLNSRL LEIFCSEIGV NHTHLLFHTE
301 VRWLSQGKVL SRVYELRNEI YIFLVEKQSH LANIFEDDIW VTKLAYLSDI
351 FGILNELSLK MQGKNNDIFQ YLEHILGFQK TLLLWQARLK SNRPSYYMFP
401 TLLQHIEENI INEDCLKEIK LEILLHLTSL SQTFNYYFPE EKFESLKENI
451 WMKDPFAFQN PESIIELNLE PEEENELLQL SSSFTLKNYY KILSLSAFWI
501 KIKDDFPLLS RKSILLLLPF TTTYLCELGF SILTRLKTKK RNRLNSAPDM
551 RVALSSCVPD WKELMNRQAH PSH
```

BLASTP hits

Entry AC004877_3 from database TREMBLNEW:
gene: "WUGSC:H_DJ0751H13.2"; product: "KIAA0543 protein"; Homo sapiens PAC clone DJ0751H13 from 7q35-qter, complete sequence.
Score = 86, P = 4.4e-03, identities = 46/179, positives = 78/179

Entry MD36211_1 from database TREMBL:
product: "Hermes transposase"; Musca domestica Hermes transposase gene, complete cds.
Score = 105, P = 3.0e-02, identities = 101/465, positives = 202/465

Alert BLASTP hits for DKFZphmcf1_1g13, frame 1

TREMBL:AB018309_1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens mRNA for KIAA0766 protein, complete cds., N = 1, Score = 300, P = 1.1e-23

>TREMBL:AB018309_1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens mRNA for KIAA0766 protein, complete cds.
Length = 607

HSPs:

Score = 300 (45.0 bits), Expect = 1.1e-23, P = 1.1e-23
Identities = 120/485 (24%), Positives = 229/485 (47%)

```
Query:    89 CMD-MVRTIFDDKSADKLRTIPLSDNTISRRICTIAKHLEAMLITRLQSGIDFAIQLDES 147
             CM+ ++R +  +     L+ + LS +   +RI +I ++L   L  R +    +++ LD+
Sbjct:   124 CMEVLLREVLPEH-VSVLQGVDLSPDITRQRILSIDRNLRNQLFNRARDFKAYSLALDDQ 182

Query:   148 TDIASCPTLLVYVRYVWQD-DFVEDLLCCLNLNSHIT-GLDLFTELENCLLGQYKLNWKH 205
               +A    LLV++R V  + +  EDLL  +NL  H + G  +   LE+  L    L+ +
Sbjct:   183 AFVAYENYLLVFIRGVGPELEVQEDLLTIINLTHHFSVGALMSAILES--LQTAGLSLQR 240

Query:   206 CKGISSDGTANMTGKHSRLTEKLLEATHNNAVWN--HC--FIHREALVSKEISPSLMDVL 261
               G+++  T  M G++S L   + E   +   WN  H   F+H E L S ++  +   ++
Sbjct:   241 MVGLTTTHTLRMIGENSGLVSYMREKAVSPNCWNVIHYSGFLHLELLSSYDVDVN--QII 298

Query:   262 KNAVKTVNFIKGSSLNSRLLEIFCSEIGVNHTHLLFHTEVR-WLSQGKVLSRVYELRNEI 320
                 + +  IK   +     +   +E     H   +    + WL +GK L ++  LR E+
Sbjct:   299 NTISEWIVLIKTRGVRRPEFQTLLTESESEHGERVNGRCLNNWLRRGKTLKLIFSLRKEM 358

Query:   321 YIFLVEKQSHLANIFEDDIWVTKLAYLSDIFGILNELSLKMQGKNNDIFQYLEHILGFQK 380
               FLV   +   + F D  W+    +L DI   L ELS +++          +HI  F+
Sbjct:   359 EAFLVSVGATTVH-FSDKQWLCDFGFLVDIMEHLRELSEELRVSKVFAAAAFDHICTFEV 417
```

```
Query:   381 TLLLWQARLKSNRPSYYMFPTLLQHIEE----NIINEDCLKEIKLEILLHLTSLSQTFNY 436
             L L+Q  ++     +    FP L + ++E     N  +E       + ++++    L + F
Sbjct:   418 KLNLFQRHIEEKNLTD--FPALREVVDELKQQNKEDEKIFDPDRYQMVI--CRLQKEFER 473

Query:   437 YFPEEKFESLKENIWM-KDPFAFQNPESIIELNLEPEEENELLQLSSSFTLKNYYKILSL 495
             +F + +F  +K+++ +  +PF F+   + I + +E        L +L ++  L N Y+I  L
Sbjct:   474 HFKDLRF--IKKDLELFSNPFNFKPEYAPISVRVE------LTKLQANTNLWNEYRIKDL 525

Query:   496 SAFWIKIK-DDFPLLSRKSILLLLPFTTTYLCELGFSILTRLKTKKRNRLNSA---PDMR 551
              F+  +  + +P++   +  +  F + +CE  FS LTR +        L            R
Sbjct:   526 GQFYAGLSAESYPIIKGVACKVASLFDSNQICEKAFSYLTRNQHTLSQPLTDEHLQALFR 585

Query:   552 VALSSCVPDWKELMNRQAHPSH 573
             VA +   P W +L+ R+ + S+
Sbjct:   586 VATTEMEPGWDDLV-RERNESN 606

 Score = 290 (43.5 bits), Expect = 1.5e-22, P = 1.5e-22
 Identities = 120/485 (24%), Positives = 228/485 (47%)

Query:    89 CMD-MVRTIFDDKSADKLRTIPLSDNTISRRICTIAKHLEAMLITRLQSGIDFAIQLDES 147
             CM+ ++R +  +     L+ + LS +    +RI +I ++L    L   R +     +++ LD+
Sbjct:   124 CMEVLLREVLPEH-VSVLQGVDLSPDITRQRILSIDRNLRNQLFNRARDFKAYSLALDDQ 182

Query:   148 TDIASCPTLLVYVRYVWQD-DFVEDLLCCLNLNSHIT-GLDLFTELENCLLGQYKLNWKH 205
             +A     LLV++R V  + +  EDLL +NL  H + G  +  LE+ L   L+ +
Sbjct:   183 AFVAYENYLLVFIRGVGPELEVQEDLLTIINLTHHFSVGALMSAILES--LQTAGLSLQR 240

Query:   206 CKGISSDGTANMTGKHSRLTEKLLEATHNNAVWNHCFIHREALVSKEISPSLMDV-LKNA 264
             G+++  T  M G++S L   + E  +  WN   IH  + E+ S  DV +
Sbjct:   241 MVGLTTTHTLRMIGENSGLVSYMREKAVSPNCWN--VIHYSGFLHLELLSSY-DVDVNQI 297

Query:   265 VKTVN----FIKGSSLNSRLLEIFCSEIGVNHTHLLFHTEVR-WLSQGKVLSRVYELRNE 319
             + T++    IK   +    +E   H   +   + WL +GK L ++ LR E
Sbjct:   298 INTISEWIVLIKTRGVRRPEFQTLLTESESEHGERVNGRCLNNWLRRGKTLKLIFSLRKE 357

Query:   320 IYIFLVEKQSHLANIFEDDIWVTKLAYLSDIFGILNELSLKMQGKNNDIFQYLEHILGFQ 379
             +  FLV  +  +  F D W+  +L DI  L ELS +++  +HI F+
Sbjct:   358 MEAFLVSVGATTVH-FSDKQWLCDFGFLVDIMEHLRELSEELRVSKVFAAAAFDHICTFE 416

Query:   380 KTLLLWQARLKSNRPSYYMFPTLLQHIEENIINEDCLKEIKL----EILLHLTSLSQTFN 435
             L L+Q  ++     +    FP L + ++E + +  ++ K+     + + L + F
Sbjct:   417 VKLNLFQRHIEEKNLTD--FPALREVVDE--LKQQNKEDEKIFDPDRYQMVICRLQKEFE 472

Query:   436 YYFPEEKFESLKENIWM-KDPFAFQNPESIIELNLEPEEENELLQLSSSFTLKNYYKILS 494
             +F + +F  +K+++ +  +PF F+  + I + +E  L +L ++ L N Y+I
Sbjct:   473 RHFKDLRF--IKKDLELFSNPFNFKPEYAPISVRVE------LTKLQANTNLWNEYRIKD 524

Query:   495 LSAFWIKIK-DDFPLLSRKSILLLLPFTTTYLCELGFSILTRLKTKKRNRLNSA---PDM 550
             L F+ + + +P++  + +  F +  +CE FS LTR +        L
Sbjct:   525 LGQFYAGLSAESYPIIKGVACKVASLFDSNQICEKAFSYLTRNQHTLSQPLTDEHLQALF 584

Query:   551 RVALSSCVPDWKELMNRQAHPSH 573
             RVA +   P W +L+ R+ + S+
Sbjct:   585 RVATTEMEPGWDDLV-RERNESN 606
```

Pedant information for DKFZphmcf1_1g13, frame 1

Report for DKFZphmcf1_1g13.1

```
[LENGTH]        573
[MW]            66276.85
[pI]            5.82
[HOMOL]         TREMBL:AB018309_1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens
mRNA for KIAA0766 protein, complete cds. 1e-18
[PROSITE]       MYRISTYL          3
[PROSITE]       CK2_PHOSPHO_SITE        10
[PROSITE]       TYR_PHOSPHO_SITE        1
[PROSITE]       PKC_PHOSPHO_SITE        9
[PROSITE]       ASN_GLYCOSYLATION       2
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY     8.90 %

SEQ     MTPESRDTTDLSPGGTQEMEGIVIVKVEEEDEEDHFQKERNKVESSPQVLSRSTTMNERA
SEG     ..........................xxxxxxx...........................
PRD     cccccccccccccccccccceeeeeeeeccccchhhhhhhhhhcccccceeeccccchhhhh

SEQ     LLSSYLVAYRVAKEKMAHTAAEKIILPACMDMVRTIFDDKSADKLRTIPLSDNTISRRIC
```

```
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhchhhhhhhhhhcccccceeeeecccchhhhhhh

SEQ     TIAKHLEAMLITRLQSGIDFAIQLDESTDIASCPTLLVYVRYVWQDDFVEDLLCCLNLNS
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhheeeccccccccccccccceeeeeeeccchhhhhhhhhhhhccce

SEQ     HITGLDLFTELENCLLGQYKLNWKHCKGISSDGTANMTGKHSRLTEKLLEATHNNAVWNH
SEG     ............................................................
PRD     eeeehhhhhhhhhhhhhhccccccccccccccceeeecccchhhhhhhhhhcccccceee

SEQ     CFIHREALVSKEISPSLMDVLKNAVKTVNFIKGSSLNSRLLEIFCSEIGVNHTHLLFHTE
SEG     ............................................................
PRD     hhhhhhhhhhhcccchhhhhhhhhhhheeecccccchhhhhhhhhhhccccchhhhhhhh

SEQ     VRWLSQGKVLSRVYELRNEIYIFLVEKQSHLANIFEDDIWVTKLAYLSDIFGILNELSLK
SEG     ............................................................
PRD     cccccccchhhhhhhhhhhhhhhhhhhchhhhhcccceeehhhhhhhhhhhhhhhhhhhh

SEQ     MQGKNNDIFQYLEHILGFQKTLLLWQARLKSNRPSYYMFPTLLQHIEENIINEDCLKEIK
SEG     .......................................................xxxxx
PRD     hhccccccchhhhhhhhhhhhhhhhhhhhccccccccccchhhhhhhhhhhhhcchhhhhhh

SEQ     LEILLHLTSLSQTFNYYFPEEKFESLKENIWMKDPFAFQNPESIIELNLEPEEENELLQL
SEG     xxxxx....................................xxxxxxxxxxxxxxxxxxx
PRD     hhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhcccccccccccceeecccchhhhhhhhh

SEQ     SSSFTLKNYYKILSLSAFWIKIKDDFPLLSRKSILLLLPFTTTYLCELGFSILTRLKTKK
SEG     xxx......................xxxxxxxxxxx........................
PRD     hhcccchhhhhhhhhhhhhcccccccccccchhhhhhhhhccceeeehhhhhhhhhhhhhh

SEQ     RNRLNSAPDMRVALSSCVPDWKELMNRQAHPSH
SEG     .................................
PRD     hcccccccccceeeccccccccchhhhhhccccc
```

Prosite for DKFZphmcf1_1g13.1

PS00001	216->220	ASN_GLYCOSYLATION	PDOC00001
PS00001	291->295	ASN_GLYCOSYLATION	PDOC00001
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00005	225->228	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00005	391->394	PKC_PHOSPHO_SITE	PDOC00005
PS00005	445->448	PKC_PHOSPHO_SITE	PDOC00005
PS00005	485->488	PKC_PHOSPHO_SITE	PDOC00005
PS00005	510->513	PKC_PHOSPHO_SITE	PDOC00005
PS00005	538->541	PKC_PHOSPHO_SITE	PDOC00005
PS00006	55->59	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	95->99	CK2_PHOSPHO_SITE	PDOC00006
PS00006	136->140	CK2_PHOSPHO_SITE	PDOC00006
PS00006	183->187	CK2_PHOSPHO_SITE	PDOC00006
PS00006	189->193	CK2_PHOSPHO_SITE	PDOC00006
PS00006	256->260	CK2_PHOSPHO_SITE	PDOC00006
PS00006	445->449	CK2_PHOSPHO_SITE	PDOC00006
PS00006	463->467	CK2_PHOSPHO_SITE	PDOC00006
PS00006	546->550	CK2_PHOSPHO_SITE	PDOC00006
PS00007	364->372	TYR_PHOSPHO_SITE	PDOC00007
PS00008	137->143	MYRISTYL	PDOC00008
PS00008	273->279	MYRISTYL	PDOC00008
PS00008	289->295	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphmcf1_1g13.1)

DKFZphtes3_14g5
----

group: testes derived

DKFZphtes3_14g5 encodes a novel 379 amino acid protein with strong similarity to murine cell growth regulating nucleolar protein LYAR.

The novel protein is very similar to murine Ly-1 antibody reactive clone protein (LYAR). It contains a ATP/GTP-binding site motif A (P-loop, interacts with one of the phosphate groups of a ATP/GTP nucleotide), but not the zinc finger motif and and nuclear localization signals of lyar.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to cell growth regulating nucleolar protein LYAR, of mouse

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1503 bp
Poly A stretch at pos. 1467, polyadenylation signal at pos. 1440

```
   1 CCCAGAGGTC CGACCTGGGA GGCTGGGGCT CAGAGAGCAA TGTTTGCTGT
  51 CTTCCATTGG AGTGACTGAA TTTCTACATG ACGGCTTTTT GACAAGACTT
 101 AAAACCTGTC TTGGATAGAG AATATTTAGC CATTTACCTA AAAATGGTAT
 151 TTTTTACATG CAATGCATGT GGTGAATCAG TGAAGAAAAT ACAAGTGGAA
 201 AAGCATGTGT CTGTTTGCAG AAACTGTGAA TGCCTTTCTT GCATTGACTG
 251 CGGTAAAGAT TTCTGGGGCG ATGACTATAA AAACCACGTG AAATGCATAA
 301 GTGAAGATCA GAAGTATGGT GGCAAAGGCT ATGAAGGTAA AACCCACAAA
 351 GGCGACATCA AACAGCAGGC GTGGATTCAG AAAATTAGTG AATTAATAAA
 401 GAGACCCAAT GTCAGCCCCA AAGTGAGAGA ACTTTTAGAG CAAATTAGTG
 451 CTTTTGACAA CGTTCCCAGG AAAAAGGCAA AATTTCAGAA TTGGATGAAG
 501 AACAGTTTAA AAGTTCATAA TGAATCCATT CTGGACCAGG TGTGGAATAT
 551 CTTTTCTGAA GCTTCCAACA GCGAACCAGT CAATAAGGAA CAGGATCAAC
 601 GGCCACTCCA CCCAGTGGCA AATCCACATG CAGAAATCTC CACCAAGGTT
 651 CCAGCCTCCA AAGTGAAAGA CGCCGTGGAA CAGCAAGGGG AGGTGAAGAA
 701 GAATAAAAGA GAAAGAAAGG AAGAACGGCA GAAGAAAAGG AAAAGAGAAA
 751 AGAAAGAACT AAAGTTAGAA AACCACCAGG AAAACTCAAG GAATCAGAAG
 801 CCTAAGAAGC GCAAAAAGGG ACAGGAGGCT GACCTTGAGG CTGGTGGGGA
 851 GGAAGTCCCT GAGGCCAATG GCTCTGCAGG GAAGAGGAGC AAGAAGAAGA
 901 AGCAGCGCAA GGACAGCGCC AGTGAGGAAG AGGCACGCGT GGGCGCAGGG
 951 AAGAGGAAGC GGAGGCACTC GGAAGTTGAA ACAGATTCTA AGAAGAAAAA
1001 GATGAAGCTC CCAGAGCATC CTGAGGGCGG AGAACCAGAA GACGATGAGG
1051 CTCCTGCAAA AGGTAAATTC AACTGGAAGG GAACTATTAA AGCAATTCTG
1101 AAACAGGCCC CAGACAATGA AATAACCATC AAAAAGCTAA GGAAAAAGGT
1151 TTTAGCTCAG TACTACACAG TGACAGATGA GCATCACAGA TCCGAAGAGG
1201 AACTCCTGGT CATCTTTAAC AAGAAAATCA GCAAGAACCC TACCTTTAAG
1251 TTATTAAAGG ACAAAGTCAA GCTTGTGAAA TGAACATTTG TGTATTTAAA
1301 AATTGAATCC ATTCTGCTGA CTTCTTCCTT TCACTGCTGT TTATAAAATG
1351 TGTAATGAAT TCTAACAACT CAAATTTTGC TTTTTGAAGC TGTATTTTTA
1401 AGTTAAGAAA ATATATTTTT GGTATAACTT TTATGAGAAA AATAAAATAT
1451 ATTCTGGTCC AAACTTCAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1501 AAA
```

BLAST Results

No BLAST result

Medline entries

93259460:
LYAR, a novel nucleolar protein with zinc finger DNA-binding motifs, is involved in cell
growth regulation.

Peptide information for frame 3

ORF from 144 bp to 1280 bp; peptide length: 379
Category: strong similarity to known protein
Classification: Cell division
Prosite motifs: ATP_GTP_A (60-68)

```
  1 MVFFTCNACG ESVKKIQVEK HVSVCRNCEC LSCIDCGKDF WGDDYKNHVK
 51 CISEDQKYGG KGYEGKTHKG DIKQQAWIQK ISELIKRPNV SPKVRELLEQ
101 ISAFDNVPRK KAKFQNWMKN SLKVHNESIL DQVWNIFSEA SNSEPVNKEQ
151 DQRPLHPVAN PHAEISTKVP ASKVKDAVEQ QGEVKKNKRE RKEERQKKRK
201 REKKELKLEN HQENSRNQKP KKRKKGQEAD LEAGGEEVPE ANGSAGKRSK
251 KKKQRKDSAS EEEARVGAGK RKRRHSEVET DSKKKKMKLP EHPEGGEPED
301 DEAPAKGKFN WKGTIKAILK QAPDNEITIK KLRKKVLAQY YTVTDEHHRS
351 EEELLVIFNK KISKNPTFKL LKDKVKLVK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14g5, frame 3

PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse, N = 1, Score = 1410, P = 2.7e-144

SWISSPROT:YQ58_CAEEL HYPOTHETICAL 28.5 KD PROTEIN C16C10.8 IN CHROMOSOME III., N = 1, Score = 381, P = 2.9e-35

TREMBL:AC003058_18 gene: "F27F23.18"; product: "putative RNA-binding protein"; Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence, complete sequence., N = 3, Score = 139, P = 4e-15

PIR:S70049 nucleic acid-binding protein YCR087c-a - yeast (Saccharomyces cerevisiae), N = 1, Score = 164, P = 1.4e-11

```
>PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse
           Length = 388

  HSPs:

 Score = 1410 (211.6 bits), Expect = 2.7e-144, P = 2.7e-144
 Identities = 275/388 (70%), Positives = 317/388 (81%)

Query:     1 MVFFTCNACGESVKKIQVEKHVSVCRNCECLSCIDCGKDFWGDDYKNHVKCISEDQKYGG 60
             MVFFTCNACGESVKKIQVEK VS CRNCECLSCIDCGKDFWGDDYK+HVKCISE QKYGG
Sbjct:     1 MVFFTCNACGESVKKIQVEKQVSNCRNCECLSCIDCGKDFWGDDYKSHVKCISEGQKYGG 60

Query:    61 KGYEGKTHKGDIKQQAWIQKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNWMKN 120
             KGYE KTHKGD KQQAWIQKI+ELIK+PNVSPKVRELL+QISAFDNVP KKAKFQNWMKN
Sbjct:    61 KGYEAKTHKGDAKQQAWIQKINELIKKPNVSPKVRELLQQISAFDNVPIKKAKFQNWMKN 120

Query:   121 SLKVHNESILDQVWNIFSEASNSEPVNKEQDQRPLHPVANPHAEIS-TKVPASKVKDAVE 179
             SLKVH++S+L+QVW+IFSEAS+SE   ++Q Q P H  A PHAE+  TKVP++K     E
Sbjct:   121 SLKVHSDSVLEQVWDIFSEASSSE---QDQQQPPSH-TAKPHAEMPITKVPSAKTNGTTE 176

Query:   180 QQGEVKKNKRERKEERQKKRKREKKELKLENHQENSRNQKPKKRKKGQEADLEAGGEEVP 239
             +Q E KKNKRERKEERQK RK+EKKELKLENHQEN R QKPKKRKK QEA  EA GE+
Sbjct:   177 EQTEAKKNKRERKEERQKNRKKEKKELKLENHQENLRGQKPKKRKKNQEAGHEAAGEDGA 236

Query:   240 EANG-------SAGKRSKKKKQRKDSASEEEA----RVGAGKRKR-RHSEVETDSKKKKM 287
             + +G         G+ S++   R     E+ A    +  AGKRKR +HS  E+  KKKKM
Sbjct:   237 DGSGPPEKKKAQGGQASEEGADRNGGPGEDRAEGQTKTAAGKRKRPKHSGAESGYKKKKM 296

Query:   288 KLPEHPEGGEPEDDEAPAKGKFNWKGTIKAILKQAPDNEITIKKLRKKVLAQYYTVTDEH 347
             KLPE PE GE +D EAP+KGKFNWKGTIKA+LKQAPDNEI++KKL+KKV+AQY+ V ++
Sbjct:   297 KLPEQPEEGEAKDHEAPSKGKFNWKGTIKAVLKQAPDNEISVKKLKKKVIAQYHAVMNDT 356

Query:   348 HRSEEELLVIFNKKISKNPTFKLLKDKVKLVK 379
                EEELL IFN+KIS+NPTFK+LKD+VKL+K
Sbjct:   357 SHHEEELLAIFNRKISRNPTFKVLKDRVKLLK 388
```

Pedant information for DKFZphtes3_14g5, frame 3

Report for DKFZphtes3_14g5.3

```
[LENGTH]       379
[MW]           43634.03
[pI]           9.59
[HOMOL]        PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse 1e-122
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YCR087c-a] 2e-11
[BLOCKS]       BL00603D Thymidine kinase cellular-type proteins
[BLOCKS]       BL00530C
[PROSITE]      ATP_GTP_A      1
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY    18.73 %

SEQ     MVFFTCNACGESVKKIQVEKHVSVCRNCECLSCIDCGKDFWGDDYKNHVKCISEDQKYGG
SEG     ............................................................
PRD     cccccccccccchhhhhhheeecccccceeecccccccccccccccceeeeeeccccccc

SEQ     KGYEGKTHKGDIKQQAWIQKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNWMKN
SEG     ............................................................
PRD     ccccccccchhhhhhhhhhhhhhhhhccchhhhhhhhhhhhcccccchhhhhhhhhhhhc

SEQ     SLKVHNESILDQVWNIFSEASNSEPVNKEQDQRPLHPVANPHAEISTKVPASKVKDAVEQ
SEG     ............................................................
PRD     cccccchhhhhhhhhhhhhhhcchhhhhhhhcccccccccccccccceeecccccchhhhh

SEQ     QGEVKKNKRERKEERQKKRKREKKELKLENHQENSRNQKPKKRKKGQEADLEAGGEEVPE
SEG     ....xxxxxxxxxxxxxxxxxxxxxxxxxx..............................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccchhhhhhchhhhhccccc

SEQ     ANGSAGKRSKKKKQRKDSASEEEARVGAGKRKRRHSEVETDSKKKKMKLPEHPEGGEPED
SEG     ..xxxxxxxxxxxxxxxxxx.............................xxxxxxxxxxx
PRD     cccccccchhhhhhhhcchhhhhhhhhccccccccccccccchhhhhhcccccccccccc

SEQ     DEAPAKGKFNWKGTIKAILKQAPDNEITIKKLRKKVLAQYYTVTDEHHRSEEELLVIFNK
SEG     xxxxx.......................................................
PRD     cccccceeeehhhhhhhhhhhccccccchhhhhhhhhhhhhhhcchhhhhhhhhhhhhhh

SEQ     KISKNPTFKLLKDKVKLVK
SEG     ........xxxxxxxxxxx
PRD     ccccccchhhhhhhhhccc
```

Prosite for DKFZphtes3_14g5.3

```
PS00017      60->68    ATP_GTP_A                PDOC00017
```

(No Pfam data available for DKFZphtes3_14g5.3)

DKFZphtes3_14h21
----

group: nucleic acid management

DKFZphtes3_14h21 encodes a novel 648 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and a ATP/GTP-binding site motif A (P-loop) and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicases

start at Bp 33 matches Kozak consensus ACNatg

Sequenced by BMFZ

Locus: unknown

Insert length: 2200 bp
Poly A stretch at pos. 2166, polyadenylation signal at pos. 2140

```
   1 CAACGACGTC GGACGCGCCC CTTCTTGGAA CAATGTCCCA CCACGGAGGA
  51 GCTCCCAAGG CCTCTACGTG GGTCGTTGCT AGTCGGCGAA GCTCGACAGT
 101 GTCCCGAGCG CCAGAGAGGA GGCCGGCGGA GGAGTTGAAT CGAACAGGTC
 151 CTGAGGGATA TAGTGTCGGC AGAGGTGGTC GCTGGAGAGG CACCTCTAGG
 201 CCCCCGGAGG CCGTGGCCGC TGGTCACGAG GAACTGCCGC TGTGTTTTGC
 251 TTTGAAGAGC CACTTTGTTG GCGCGGTAAT CGGTCGTGGT GGGTCAAAAA
 301 TAAAGAATAT ACAAAGTACA ACAAACACCA CAATCCAAAT AATACAAGAA
 351 CAACCAGAAT CATTAGTCAA AATTTTTGGC AGCAAGGCAA TGCAAACGAA
 401 AGCAAAAGCA GTGATAGACA ATTTTGTTAA AAAGCTAGAA GAAAATTACA
 451 ATTCAGAATG CGGAATTGAT ACTGCATTCC AACCTTCTGT TGGAAAAGAT
 501 GGAAGCACAG ATAACAATGT TGTTGCAGGA GATCGGCCAT TGATAGATTG
 551 GGATCAAATT AGAGAGGAAG GTTTGAAATG GCAAAAAACA AAGTGGGCAG
 601 ATTTACCACC AATTAAGAAA AACTTTTATA AAGAGTCCAC TGCCACAAGT
 651 GCCATGTCAA AAGTAGAAGC AGATAGTTGG AGGAAAGAAA ATTTTAATAT
 701 AACGTGGGAT GACTTGAAGG ATGGGGAGAA ACGACCTATC CCCAATCCTA
 751 CCTGCACATT TGATGACGCC TTTCAATGTT ATCCTGAGGT TATGGAAAAC
 801 ATTAAAAAGG CAGGTTTTCA AAAGCCAACA CCTATTCAGT CACAGGCATG
 851 GCCCATTGTG TTGCAAGGAA TAGATCTTAT AGGAGTAGCC CAGACTGGAA
 901 CAGGAAAGAC ATTGTGTTAT TTAATGCCTG GATTTATTCA TCTGGTCCTT
 951 CAACCCAGCC TTAAAGGTCA AAGGAATAGA CCCGGCATGT TAGTTCTAAC
1001 TCCCACTCGG GAATTAGCAC TTCAAGTAGA AGGAGAATGT TGCAAATATT
1051 CATATAAAGG GCTTCGGAGT GTTTGTGTAT ATGGTGGTGG AAATAGAGAT
1101 GAACAAATAG AAGAGCTTAA AAAAGGTGTA GATATCATAA TTGCAACTCC
1151 CGGAAGATTG AATGATCTGC AAATGAGTAA CTTCGTCAAT CTGAAGAATA
1201 TAACCTACTT GGTTTTAGAT GAAGCAGACA AGATGTTGGA CATGGGATTT
1251 GAACCCCAGA TAATGAAGAT TTTGTTAGAT GTGCGCCCAG ATAGGCAGAC
1301 AGTTATGACC AGTGCTACAT GGCCTCATTC AGTTCATCGC CTCGCACAAT
1351 CTTATTTGAA AGAACCAATG ATTGTCTATG TTGGTACATT GGATCTAGTT
1401 GCTGTAAGTT CAGTGAAGCA AAATATAATT GTAACCACCG AGGAAGAGAA
1451 ATGGAGTCAC ATGCAAACTT TTCTACAGAG TATGTCATCC ACAGACAAAG
1501 TCATTGTCTT CGTTTCTCGA AAAGCTGTTG CGGATCACTT ATCAAGTGAC
1551 CTAATACTTG GAAATATATC AGTAGAGTCT CTGCATGGAG ATAGAGAACA
1601 GAGAGATCGG GAGAAAGCAT TAGAGAACTT TAAAACAGGC AAAGTGAGAA
1651 TACTAATTGC AACTGATCTA GCCTCTAGAG GACTTGATGT CCATGACGTT
1701 ACACATGTCT ATAATTTTGA CTTTCCACGG AATATTGAAG AATACGTACA
1751 CCGAATAGGG CGCACGGGAA GAGCAGGGAG GACTGGTGTT TCCATTACAA
1801 CTTTGACTAG AAATGATTGG AGGGTTGCCT CTGAATTGAT TAATATTCTG
1851 GAAAGAGCAA ATCAGAGTAT TCCAGAGGAG CTTGTATCAA TGGCTGAGAG
1901 GTTTGAGGCA CATCAACGGA AAAGGGAAAT GGAAAGAAAA ATGGAAAGAC
1951 CTCAAGGAAG GCCCAAGAAG TTTCATTAAT GTCTTCTGTA CTAGTGGGGT
2001 AGAGAATTCA AGATTTTTTA GAAATATAGT AAGACAGAAG TATTGGACAT
2051 GTTGGCAGTA TGAAGAGACC GGACTGATTT GACTGATTCT TAAAATAATA
2101 GTGTTTGAAA ATATAGAATC CAGTGTTTTA TACTTTCTTT AATAAAAATA
2151 GAAGTATTTA AACTTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 33 bp to 1976 bp; peptide length: 648
Category: strong similarity to known protein
Classification: Nucleic acid management
Prosite motifs: ATP_GTP_A (286-294)
DEAD_ATP_HELICASE (394-403)

```
  1 MSHHGGAPKA STWVVASRRS STVSRAPERR PAEELNRTGP EGYSVGRGGR
 51 WRGTSRPPEA VAAGHEELPL CFALKSHFVG AVIGRGGSKI KNIQSTTNTT
101 IQIIQEQPES LVKIFGSKAM QTKAKAVIDN FVKKLEENYN SECGIDTAFQ
151 PSVGKDGSTD NNVVAGDRPL IDWDQIREEG LKWQKTKWAD LPPIKKNFYK
201 ESTATSAMSK VEADSWRKEN FNITWDDLKD GEKRPIPNPT CTFDDAFQCY
251 PEVMENIKKA GFQKPTPIQS QAWPIVLQGI DLIGVAQTGT GKTLCYLMPG
301 FIHLVLQPSL KGQRNRPGML VLTPTRELAL QVEGECCKYS YKGLRSVCVY
351 GGGNRDEQIE ELKKGVDIII ATPGRLNDLQ MSNFVNLKNI TYLVLDEADK
401 MLDMGFEPQI MKILLDVRPD RQTVMTSATW PHSVHRLAQS YLKEPMIVYV
451 GTLDLVAVSS VKQNIIVTTE EEKWSHMQTF LQSMSSTDKV IVFVSRKAVA
501 DHLSSDLILG NISVESLHGD REQRDREKAL ENFKTGKVRI LIATDLASRG
551 LDVHDVTHVY NFDFPRNIEE YVHRIGRTGR AGRTGVSITT LTRNDWRVAS
601 ELINILERAN QSIPEELVSM AERFEAHQRK REMERKMERP QGRPKKFH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14h21, frame 3

TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid Y54G11A, N = 1, Score = 1008, P = 1.1e-101

TREMBL:SPBP8B7_16 gene: "dbp2"; "SPBP8B7.16c"; product: "p68-like protein."; S.pombe chromosome II pl p8B7., N = 1, Score = 971, P = 9.1e-98

PIR:S13757 RNA helicase DBP2 - yeast (Saccharomyces cerevisiae), N = 1, Score = 970, P = 1.2e-97

PIR:S14048 RNA helicase dbp2 - fission yeast (Schizosaccharomyces pombe), N = 1, Score = 961, P = 1e-96

PIR:A57514 RNA helicase HEL117 - rat, N = 2, Score = 888, P = 7.8e-91

>TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid Y54G11A
Length = 504

HSPs:

Score = 1008 (151.2 bits), Expect = 1.1e-101, P = 1.1e-101
Identities = 211/473 (44%), Positives = 298/473 (63%)

```
Query:   174 DQIREEGLKWQKTKWADLPPIKKNFYKESTATSAMSKVEADSWRKENFNITWDDLKDGEK 233
             D++++E   W K        PI ++ YK        +S  +              +  ++
Sbjct:    23 DRLKDENFSWMK-------PIVRDLYKIPNEQKNLSPEQLQELYTNGGVMKVYPFREEST 75

Query:   234 RPIPNPTCTFDDAFQCYPEVMENIKKAGFQKPTPIQSQAWPIVLQGIDLIGVAQTGTGKT 293
               IP P  +F+ AF     +M  I+K GF+KP+PIQSQ WP++L G D IGV+QTG+GKT
Sbjct:    76 VKIPPPVNSFEQAFGSNASIMGEIRKNGFEKPSPIQSQMWPLLLSGQDCIGVSQTGSGKT 135

Query:   294 LCYLMPGFIHLVLQPSL-----KGQRNRPGMLVLTPTRELALQVEGECCKYSYKGLRSVC 348
             L +L+P  +H+  Q +      + Q+  P +LVL+PTRELA Q+EGE  KYSY G +SVC
Sbjct:   136 LAFLLPALLHIDAQLAQYEKNDEEQKPSPFVLVLSPTRELAQQIEGEVKKYSYNGYKSVC 195

Query:   349 VYGGGNRDEQIEELKKGVDIIIATPGRLNDLQMSNFVNLKNITYLVLDEADKMLDMGFEP 408
```

```
             +YGGG+R EQ+E  + GV+I+IATPGRL DL      ++L ++TY+VLDEAD+MLDMGFE
Sbjct:   196 LYGGGSRPEQVEACRGGVEIVIATPGRLTDLSNDGVISLASVTYVVLDEADRMLDMGFEV 255

Query:   409 QIMKILLDVRPDRQTVMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVT 468
             I +IL ++RPDR    +TSATWP  V +L   Y KE ++    G+LDL +  SV Q
Sbjct:   256 AIRRILFEIRPDRLVALTSATWPEGVRKLTDKYTKEAVMAVNGSLDLTSCKSVTQFFEFV 315

Query:   469 TEEEKW---SHMQTFLQSMSSTD-KVIVFVSRKAVADHLSSDLILGNISVESLHGDREQR 524
              + ++       +  FL +     + K+I+FV  K +ADHLSSD  +  I+ + LHG R Q
Sbjct:   316 PHDSRFLRVCEIVNFLTAAHGQNYKMIIFVKSKVMADHLSSDFCMKGINSQGLHGGRSQS 375

Query:   525 DREKALENFKTGKVRILIATDLASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAGRT 584
             DRE +L    ++G+V+IL+ATDLASRG+DV D+THV N+DFP +IEEYVHR+GRTGRAGR
Sbjct:   376 DREMSLNMLRSGEVQILVATDLASRGIDVPDITHVLNYDFPMDIEEYVHRVGRTGRAGRK 435

Query:   585 GVSITTLTRNDWRVASELINILERANQSIPEELVSMAERFEAHQRKREMERKMERPQGRP 644
             G +++ L  ND       LI ILE++ Q +P++L   AE++      K +  R   RP  R
Sbjct:   436 GEAMSFLWWNDRSNFEGLIQILEKSEQEVPDQLRRDAEKYRL---KCQSGRDGPRPSFRN 492

Query:   645 KK 646
              K
Sbjct:   493 NK 494
```

Pedant information for DKFZphtes3_14h21, frame 3

Report for DKFZphtes3_14h21.3

```
[LENGTH]        648
[MW]            72873.51
[pI]            8.84
[HOMOL]         TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid Y54G11A 1e-
101
[FUNCAT]        04.01.04 rrna processing      [S. cerevisiae, YNL112w] 2e-97
[FUNCAT]        30.10 nuclear organization    [S. cerevisiae, YNL112w] 2e-97
[FUNCAT]        04.05.03 mrna processing (splicing)   [S. cerevisiae, YPL119c] 4e-72
[FUNCAT]        30.03 organization of cytoplasm      [S. cerevisiae, YOR204w] 2e-70
[FUNCAT]        05.04 translation (initiation, elongation and termination)  [S. cerevisiae,
YOR204w] 2e-70
[FUNCAT]        06.10 assembly of protein complexes   [S. cerevisiae, YBR237w] 1e-61
[FUNCAT]        l genome replication, transcription, recombination and repair     [H.
influenzae, HI0892] 2e-49
[FUNCAT]        j mrna translation and ribosome biogenesis   [H. influenzae, HI0231 RNA] 1e-48
[FUNCAT]        04.99 other transcription activities [S. cerevisiae, YDL160c] 9e-45
[FUNCAT]        04.05.01.07 chromatin modification    [S. cerevisiae, YMR290c] 3e-44
[FUNCAT]        09.01 biogenesis of cell wall         [S. cerevisiae, YJL033w] 2e-36
[FUNCAT]        98 classification not yet clear-cut   [S. cerevisiae, YOR046c] 7e-32
[FUNCAT]        30.16 mitochondrial organization      [S. cerevisiae, YDR194c] 2e-28
[FUNCAT]        99 unclassified proteins      [S. cerevisiae, YGL064c] 5e-10
[FUNCAT]        11.10 cell death       [S. cerevisiae, YMR190c] 2e-08
[FUNCAT]        03.19 recombination and dna repair    [S. cerevisiae, YMR190c] 2e-08
[FUNCAT]        r general function prediction         [M. jannaschii, MJ1401] 1e-07
[BLOCKS]        BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]        BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]        BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]        BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW]         nucleus 4e-96
[PIRKW]         RNA binding 3e-87
[PIRKW]         DEAD box 5e-50
[PIRKW]         transmembrane protein 4e-27
[PIRKW]         DNA binding 3e-67
[PIRKW]         recF recombination pathway 3e-10
[PIRKW]         ATP 4e-96
[PIRKW]         purine nucleotide binding 5e-50
[PIRKW]         P-loop 4e-96
[PIRKW]         hydrolase 9e-45
[PIRKW]         protein biosynthesis 5e-50
[PIRKW]         ATP binding 1e-61
[SUPFAM]        WW repeat homology 8e-88
[SUPFAM]        DEAD/H box helicase homology 4e-96
[SUPFAM]        unassigned DEAD/H box helicases 7e-87
[SUPFAM]        ATP-dependent RNA helicase DBP1 4e-96
[SUPFAM]        ATP-dependent RNA helicase DHH1 2e-43
[SUPFAM]        recQ protein 3e-10
[SUPFAM]        Bloom's syndrome helicase 5e-07
[SUPFAM]        translation initiation factor eIF-4A 5e-50
[SUPFAM]        recQ helicase homology 3e-10
[SUPFAM]        tobacco ATP-dependent RNA helicase DB10 8e-88
[PROSITE]       DEAD_ATP_HELICASE      1
```

```
[PROSITE]       ATP_GTP_A      1
[PFAM]          Helicases conserved C-terminal domain
[PFAM]          KH domain family of RNA binding proteins
[PFAM]          DEAD and DEAH box helicases
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY     8.49 %

SEQ    MSHHGGAPKASTWVVASRRSSTVSRAPERRPAEELNRTGPEGYSVGRGGRWRGTSRPPEA
SEG    .........xxxxxxxxxxxxxxxxx..................................
PRD    cccccccccceeeeeecccccccccccccccccccccccccccccccccccccccccccc

SEQ    VAAGHEELPLCFALKSHFVGAVIGRGGSKIKNIQSTTNTTIQIIQEQPESLVKIFGSKAM
SEG    ...............................xxxxxxxxxxxxxxx..............
PRD    cccccccccchhhhhcccceeeecccccccccccccccceeeeecccccceeeeeccchh

SEQ    QTKAKAVIDNFVKKLEENYNSECGIDTAFQPSVGKDGSTDNNVVAGDRPLIDWDQIREEG
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccc

SEQ    LKWQKTKWADLPPIKKNFYKESTATSAMSKVEADSWRKENFNITWDDLKDGEKRPIPNPT
SEG    ............................................................
PRD    chhhhhhcccccccccccccccccchhhhhhhhhhhhhhheeeeeccccccccccccccc

SEQ    CTFDDAFQCYPEVMENIKKAGFQKPTPIQSQAWPIVLQGIDLIGVAQTGTGKTLCYLMPG
SEG    ............................................................
PRD    cccccccccchhhhhhhhhccccccccccccccccccccceeeeeeeccccccceeeecce

SEQ    FIHLVLQPSLKGQRNRPGMLVLTPTRELALQVEGECCKYSYKGLRSVCVYGGGNRDEQIE
SEG    ............................................................
PRD    eeeecccccccccccccceeeeeccchhhhhhhhhhhhhhhccceeeeeeecccccchhhh

SEQ    ELKKGVDIIIATPGRLNDLQMSNFVNLKNITYLVLDEADKMLDMGFEPQIMKILLDVRPD
SEG    ............................................................
PRD    hhhhceeeeeecccccchhhhhhhcccccccceeeehhhhhhhhhccccchhhhhhhhhhccc

SEQ    RQTVMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVTTEEEKWSHMQTF
SEG    ............................................................
PRD    ceeeeeecccchhhhhhhhhhhheeeeeecccccccccccccceeehhhhhchhhhhhhhh

SEQ    LQSMSSTDKVIVFVSRKAVADHLSSDLILGNISVESLHGDREQRDREKALENFKTGKVRI
SEG    ............................................................
PRD    hhhhcccceeeeeeehhhhhhhhhhhhhccccceeeccccccchhhhhhhhhhhhhhccccee

SEQ    LIATDLASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAGRTGVSITTLTRNDWRVAS
SEG    ...............................xxxxxxxxxxxx.................
PRD    eeehhhhhhcccccceeeeeeecccccccccceeeeccccccccceeeeeeeccccchhhh

SEQ    ELINILERANQSIPEELVSMAERFEAHQRKREMERKMERPQGRPKKFH
SEG    ............................xxxxxxxxxxx.........
PRD    hhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhcccccccccc
```

Prosite for DKFZphtes3_14h21.3

```
PS00017    286->294    ATP_GTP_A            PDOC00017
PS00039    394->403    DEAD_ATP_HELICASE    PDOC00039
```

Pfam for DKFZphtes3_14h21.3

```
HMM_NAME      DEAD and DEAH box helicases

HMM                *gLpPWILRnIyeMGFEkPTPIQQqAIPiILeGRDVMACAQTGSGKTAAF
                       P++++NI+++GF KPTPIQ+QA+PI+L+G D+++ AQTG+GKT+++
Query         248   QCYPEVMENIKKAGFQKPTPIQSQAWPIVLQGIDLIGVAQTGTGKTLCY    296

HMM                lIPMLQHIDwdPWpqpPQd..PrALILAPTRELAMQIQEEcRkFgkHMng
                   L+P ++H+   +P +++ Q+   P +L+L+PTRELA+Q++ EC K+++  +
Query         297  LMPGFIHLVLQP-SLKGQRNRPGMLVLTPTRELALQVEGECCKYSYK-G-    343

HMM                IRImcIYGGtnMRdQMRmLeRGpPHIVIATPGRLIDHIERgtldLDrIeM
                   +R++C+YGG N ++Q+++L++G+ +I+IATPGRL D+ +++ ++L++I++
Query         344  LRSVCVYGGGNRDEQIEELKKGV-DIIIATPGRLNDLQMSNFVNLKNITY    392

HMM                LVMDEADRMLDMGFIDQIRrIMrqIPMpwNRQTMMFSATMPdeIqELARr
                   LV+DEAD+MLDMGF++QI++I+   ++  ++RQT+M SAT+P ++ +LA
Query         393  LVLDEADKMLDMGFEPQIMKILLDVR--PDRQTVMTSATWPHSVHRLAQS    440
```

```
HMM            FMRNPIRInId.MdElTtnEnIkQwYiyVerEMWKfdcLcrLIe*
               ++++P  + ++  D  +++  +KQ +I+   E++K +  ++++
Query      441 YLKEPMIVYVGTLDLVAVS-SVKQNIIVTT-EEEKWSHMQTFLQ    482


HMM_NAME   KH domain family of RNA binding proteins

HMM            *rIiIPedhMGMIIGKGGsNIRqIREEYgvrINIPdecCeDstdRIITIt
                + + ++++G++IG+GGS I++I++ ++++I I++E+ +    +  + I
Query       71  CFALKSHFVGAVIGRGGSKIKNIQSTTNTTIQIIQEQ-P---ESLVKIF   115

HMM            G*
               G
Query      116 G    116


HMM_NAME   Helicases conserved C-terminal domain

HMM            *EileeWLknl....GIrvmYIHGdMpQeERdeIMddFNnGEynVLIcTD
                + +++ L+ +     +I+V ++HGD++Q++R++++++F++G+ ++LI+TD
Query      497  KAVADHLSSDLILGNISVESLHGDREQRDREKALENFKTGKVRILIATD   545

HMM            VggRGIDIPdVNHVINYDMPWNPEqYIQRIGRTgRIG*
               +++RG+D+ DV HV+N+D+P+N+E Y++RIGRTGR+G
Query      546 LASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAG    582
```

DKFZphtes3_14p14
----

group: testes derived

DKFZphtes3_14p14 encodes a novel 159 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3969 bp
Poly A stretch at pos. 3948, polyadenylation signal at pos. 3927

```
   1 GAAGCCCAGG CTCTCCTTAG TTGACTGTGT GTTAATCACC CAGCAATTTC
  51 ATTACTCAAC AGCTCTCCAG AGTTGCACAT TACAGCTGGG GTAGAAATTG
 101 GGTGCTGAAG GCCAGGCAGA GCATTTGGCT GTAGGGAGGC CGATCCTCCT
 151 CGGGCCTGTT ACCGGCGGGT CTTTGTTCTT AGACCTGGGG TTCTTGGCCT
 201 CACGGATTCC AAGGAATGGA ACGTTGGGCC ATGCGTGTGA ACGAGCTCTA
 251 TGTCGATGAC CCAGACAAGG ACAGCGGTGG CAAGATCGAC GTCAGTCTGA
 301 ACATCAGTTT ACCCAATCTG CACTGCGAGT TGGTTGGGCT TGACATTCAG
 351 GATGAGATGG GCAGGCACGA AGTGGGCCAC ATCGACAACT CCATGAAGAT
 401 CCCGCTGAAC AATGGGGCAG GCTGCCGCTT CGAGGGGCAG TTCAGCATCA
 451 ACAAGGTATG GAAGCCCTGC CTCAGCCCTT TCTACCTGCT CCCCTTTCCT
 501 GCTGTCTCCC CGCTCCCTGG AAACTGGTTG TGGAGGCACT CACTCGACCT
 551 GACCCTGACA CAGCCCCCAG CAAGCGAGGG TTCGTGTCCA GCTGCCTGGC
 601 CGTTCCTGCT GAGAATCTGG ATGGGGGTCC AGGCTCCCTG GGGTTTTAAG
 651 CCCCTGATGG CTGGTTCAGG AAGGAGCTAC TCTTCTCTCC AGTGAGGGGG
 701 ACAATGATGA GAAGACCTGA GGATTTGCAG CCCCCAGCCC TGGGTTCAAG
 751 TCCCAGCTCT ACCCCTTCTT GGCCCCTACA AGTCACTTGA CCCATCTTAG
 801 GCTGAGGGTG TGATGGCGAT AATAGTATCA CGATACCACC CACTTCACAA
 851 AGTTTGTGTG GGGATTAAAT GAGCTAATGC AGATTCATTC ATTCAGAAAA
 901 ATTTTTGAAT GGCACGTTCT GTGTTCCAGG GTCGGTGATA GGCTCTGGGG
 951 CAGCGTTCCT GGGCTGGTGG GGCTCCCATT CTGGTAGAGG GAGACAGTCT
1001 ACAAACCAGA AAGCATCAGG GATGCTAAGT GCAGTGATGA GGAATAAAGC
1051 CAAGGGGAGT GAGATGAGGT GGGCTTGAAA GTACCTTGTC CGCTCAGAAG
1101 GACCATTCAA GGTTCACTGT TGTTTTGTCC TCAGAACCAG GAGCTTCAGA
1151 TCCTAAGTCA AGTGGGTGAA CGCAGTGCCC TTGGGAGGGC CGAGGCACCC
1201 GGTGGCAGCT GGCAGGGTTT TGCTCAGCAC GTGCCGGCCT TCCTCGAAGC
1251 TCGGTACTGT CACAGTGGAG CCTCTCAACA ACGCTGTGAG GCAGCACCAT
1301 TTGACAGGTT AGGATGCTGG GGCCCAGAGA GGTTAAGTGT CTTGCCCGAG
1351 GTCACACAGC TATCTGCATG TCCCACAACT CCCCTTCCCA GCCCCAGCCA
1401 AACTGAGCCA CTGGCCACTC CTGGCTTCTC CTTGTCCCTC CTGCAGCCTC
1451 TGCTCAGAAC GCCCTTCCTC CAGACCCTGA CACCTGAGCT GGGGTTGCAA
1501 AGTCACTGGC CACATCCAGC CCAAAGATAA ATTTTGTTTG TCCAGTATAG
1551 CATTTAACTG CATCAGAACC AGTATGAAAA GACCAGGAAT CCAGATTTCT
1601 GGCTTTTAAA AGTCAGAGGC TCTCACTACA CTGGGTCCGT GTTCCCGCTA
1651 TGACAATGAC CTGGCACCAA TGGGCAGTGT TCCCCTTTAG AGAGGGTGTG
1701 TGCTGTCCCT TCCCACAGTC CCTGGCAGGC GGCTGGAAGG CCAGGCCTGG
1751 TCATCTGTCA AGCAGGGTGG ACTTCTTACG TGACAGTTCA GGGCTCCCTT
1801 AAGTGCTAAA GCAGAAGCTG CAAGGCTTTC TTAAGGTTTC GAGTGTTGCT
1851 GGGAGAAATC TGCTGCATGT TGTGGGTTAA AGGGAGTCTC TCACCAGCCC
1901 AGGCCCTCAG GAGGAGGAGA TACCAGGAGG CAGGGATGCT GGGGGTCGTG
1951 GTTCACTGGG GGCTCTCTCT GCCCATGAGC TGCCACACAG CACCTTTGCC
2001 ATGCCCCGTA ATTTGGATTT TATGGTGGTT GTGATGGAAA GCCATTTGAG
2051 GGTTTTGAAC AGGGAGGCAA TGTAATCAGA TTTATGCCTT AGAACTGGAC
2101 TATCCAATAG GTTGCCACCA GCCACATAAG GCTATTTAAA TTAATTCAAA
2151 TTAAATGTAC AATTCAGTCA CTCATTCTCA TCAACCACAT TTCAAGTGCT
2201 CAAAGCCACG TGCTGGCTAG GGGCCACAGC GTTAGACAGT GCAGAGAGAA
2251 AGCACTTCCA TCGCTGAGGA AAGTTCTGCT GGACCGCACA CCCTTAGAAG
2301 GATGGCTCTG GTGGCCGGGC GCGGTGGCTC AAACCTGTAA TCCCAGCACT
2351 TTGGGAGGCC GAGGTGGGTG GATCACGAGG TCAGGAGATC GAGACCATCC
2401 CGGCTAACAT GGTGAAACCC TGCCTCTACT AAAAATACAA AAAAAAACAA
2451 AATTAGCCGG GCGTGGTTGC GGGCACCTGT AGTCCCAGCT ACTCAGGAGG
2501 CTGAGGCGGG AGAATGGCAT GAACCCGGGA GGTGGAGCTT GCAGTGAGCC
2551 AAGATCGTAC CACTGCACTC CAGTCTGGGC GACAGAGTGA GACTCCATCT
2601 CAAAACAAAC AAAAAAAGGA TGGGGCTGGG CTGGAGAGGG TGGCAGGCAG
2651 TGGTTGTGGC AGTGGAGCTG GGGAGATGTG GTCGGATTAG GGAGGTAGAA
2701 TCAATAAGAC TCAGTGAAGA ATCGGATGTG GGGGTAAGGG CACATGTGGA
```

```
2751 AGCAAAGAAA CCTTTGACGT CTTTGTCTTG ACAACCGGGT GGTCCTGTTT
2801 CTAGACATGG AAGCTTAGAA AAGCCTGGAG TCTGTGGGAA GTAGGTAGGG
2851 CTGGGCACTG GTCATTCCAC TCTGGTTTCC TTTGGGGTTC CCATTAGGTG
2901 TCTACAGGGA GAGGTGAAAT TGGAAGTTGG AGGTGTGGAG AGTTCAGGAG
2951 AGGGTTCTGG ACCACAGATG TTGAGGTGGG AGTCATTAGT GAATAGATGA
3001 TGTTGGAAGT CATGGGTCCT CAGAGTGGGG GCTCCTTAAG CCTCCAGGCC
3051 AGCAGCATCA GCATCACCTG GGAGATTGTT AGGAATGCAG ATTCTCAGGC
3101 CCCCCTAAGA CCCACCGACT CTGTGCTAGA ACAAGCGCCC CTCAGAGATT
3151 CTGATGCCAC TGAAGTTTGA GGAGCATTGG TTTAAGCAAG ATTACCTACG
3201 GAGAGGCTGT AGATCCGTGT TCTAAACCTG GGGTCCACAG ACACCCCCAA
3251 GAAGAGCGGA TTGAATGCAA GAGATCTATG AAGTTGGATG GGGGAAAAAT
3301 TGACATCTTT ATTTTTGCTA AACTCGATCT AAAGTTTAGC ATTTCCATCT
3351 GCGATGAATG TAGGCCACAA ACCACAGTAG TATTAGCAGT GCCTGGGACC
3401 TCCTCAACAA CAGAAATTGC CGGTATTTAT AGCACGTTAC AGTTGTTGCA
3451 GATAATTTCC AGAGACTGTT TATATGCACC ACTGTTTTAA AATTACGGTG
3501 ATTGGCCAGG TGCAGTGGCT CACACCTGTA ATCCCAGCAC TTTGGGAGGC
3551 CAAAGTGGGT GGATCACTTG AGGAGTTCAA GACCAGCCTG GTCAACATGT
3601 CAAAACCCTG TATCTACAAA AAAATACAAA AGTTAACCAA GCCTATGCTT
3651 GTAGTCACAG CTACTCGGGA GGCCGAGGTG GGAGGGTCTT CTGAGCCCAG
3701 GGAGGTAGAG GCTTCAGTGA GCTGAGATCG CACCACCACA CTCCAGCCTG
3751 GGTGACAGAG TGAAACCCTT AATCAATCAG TCAATAAAAA TTACAGTAAT
3801 TATTAGACCC ACCACTAGGT CATCTTATTT GATGCATCAG TAAAGCAGCA
3851 TATTCAAATG TGGATTTTTA AATATTTTAA TTACTATTTA AATATCTCTT
3901 TACTTTGTAA TCCTATGCAT TTTACGCATT AAAACATTTT AAGCATTTAA
3951 AAAAAAAAAA AAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 216 bp to 692 bp; peptide length: 159
Category: putative protein
Classification: no clue

```
  1 MERWAMRVNE LYVDDPDKDS GGKIDVSLNI SLPNLHCELV GLDIQDEMGR
 51 HEVGHIDNSM KIPLNNGAGC RFEGQFSINK VWKPCLSPFY LLPFPAVSPL
101 PGNWLWRHSL DLTLTQPPAS EGSCPAAWPF LLRIWMGVQA PWGFKPLMAG
151 SGRSYSSLQ
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14p14, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_14p14, frame 3

Report for DKFZphtes3_14p14.3

[LENGTH]	159	
[MW]	17778.55	
[pI]	5.74	
[FUNCAT]	99 unclassified proteins	[S. cerevisiae, YAL042w] 5e-04
[KW]	Alpha_Beta	

```
SEQ  MERWAMRVNELYVDDPDKDSGGKIDVSLNISLPNLHCELVGLDIQDEMGRHEVGHIDNSM
PRD  ccchhhhhhhccccccccccceeeeeeccccccccccceeeehhhhhhcccceeecccccc

SEQ  KIPLNNGAGCRFEGQFSINKVWKPCLSPFYLLPFPAVSPLPGNWLWRHSLDLTLTQPPAS
PRD  eeeccccccceeeccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ     EGSCPAAWPFLLRIWMGVQAPWGFKPLMAGSGRSYSSLQ
PRD     cccccccchhhhhhhhhhccccccccccccccccccccc
```

(No Prosite data available for DKFZphtes3_14p14.3)

(No Pfam data available for DKFZphtes3_14p14.3)

DKFZphtes3_14p7
----

group: testes derived

DKFZphtes3_14p7 encodes a novel 702 amino acid protein with very weak similarity to kinesin associated protein KAP3.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to kinesin associated protein KAP3

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 2497 bp
Poly A stretch at pos. 2424, polyadenylation signal at pos. 2400

```
   1 GGAATCCAAA GAAACAGTTA TGATGGGGGA CTCTATGGTG AAAATAAATG
  51 GGATTTATTT AACAAAATCA AATGCTATTT GCCACTTAAA GAGTCACCCA
 101 CTTCAGCTAA CTGATGATGG AGGCTTCAGT GAAATAAAGG AGCAAGAAAT
 151 GTTCAAAGGA ACAACATCTT TACCATCTCA TCTCAAGAAT GGAGGGGACC
 201 AGGGGAAGAG ACATGCGAGG GCCTCATCAT GCCCCAGTAG CTCAGACCTG
 251 AGCAGGCTGC AAACCAAAGC AGTCCCAAAA GCTGACCTGC AAGAAGAGGA
 301 CGCAGAAATA GAAGTAGACG AAGTCTTTTG GAATACAAGG ATTGTACCGA
 351 TTTTGCGTGA ATTAGAAAAG GAAGAAAACA TTGAAACGGT TTGTGCTGCT
 401 TGCACACAAC TTCATCATGC TTTAGAGGAA GGAAACATGC TTGGAAATAA
 451 ATTTAAGGGA AGAAGTATTC TCCTGAAGAC CCTGTGTAAA CTAGTTGATG
 501 TTGGTTCAGA CTCGCTCAGC CTTAAACTTG CAAAAATAAT TCTAGCACTT
 551 AAAGTGAGTA GAAAGAATCT TCTTAATGTC TGCAAACTTA TATTTAAAAT
 601 TAGCAGGAAT GAGAAGAATG ATTCTTTGAT TCAAAATGAC AGCATTCTGG
 651 AATCATTATT GGAGGTACTA AGAAGTGAAG ACCTGCAAAC TAACATGGAA
 701 GCTTTTTTAT ACTGTATGGG GTCTATAAAG TTCATTTCTG GAAATCTGGG
 751 ATTTCTTAAT GAAATGATCA GCAAAGGTGC TGTGGAAATA CTGATAAATT
 801 TGATAAAACA AATAAATGAG AACATCAAGA AATGTGGTAC ATTTTTGCCT
 851 AATTCGGGCC ACTTGCTAGT CCAGGTGACT GCTACATTGA GAAACTTGGT
 901 TGATTCATCA TTAGTAAGAA GTAAGTTCCT AAACATCAGT GCCCTTCCCC
 951 AGCTCTGCAC GGCAATGGAA CAGTACAAGG GTGACAAGGA CGTCTGTACC
1001 AATATTGCCA GAATATTCAG CAAACTTACT TCTTACCGTG ACTGCTGCAC
1051 AGCCTTGGCC AGCTATTCCA GATGTTATGC CTTATTTCTG AATCTAATTA
1101 ACAAATACCA GAAGAAGCAG GATTTAGTCG TCCGTGTTGT TTTTATTCTT
1151 GGCAACCTGA CGGCAAAAAA TAACCAGGCT CGTGAACAAT TTTCCAAAGA
1201 GAAAGGGAGC ATCCAAACTC TGCTGTCATT ATTCCAGACG TTCCATCAGC
1251 TGGATCTGCA TTCCCAGAAG CCGGTGGGCC AACGAGGCGA GCAGCACAGG
1301 GCGCAGAGGC CGCCGTCAGA GGCAGAGGAC GTGCTCATCA AGCTGACTCG
1351 TGTGCTGGCC AACATTGCCA TCCACCCGGG CGTGGGCCCG GTGCTGGCCG
1401 CCAACCCGGG GATAGTGGGC CTGCTCCTGA CCACGCTGGA ATACAAGTCA
1451 CTTGATGATT GTGAGGAGCT GGTGATCAAT GCTACAGCGA CAATCAACAA
1501 TTTATCTTAC TACCAAGTGA AGAATTCCAT AATTCAAGAC AAAAAGCTAT
1551 ATATTGCTGA ATTGCTCTTA AAGCTTCTTG TCAGTAACAA CATGGATGGA
1601 ATCCTGGAGG CTGTGCGTGT TTTCGGAAAT CTCTCCCAGG ACCATGATGT
1651 CTGCGATTTC ATTGTGCAGA ACAATGTCCA CAGGTTCATG ATGGCGCTGC
1701 TGGATGCTCA GCATCAGGAT ATCTGCTTTT CTGCCTGTGG TGTTCTCCTC
1751 AATCTCACTG TGGATAAAGA CAAGCGTGTC ATCTTGAAAG AAGGAGGTGG
1801 CATTAAAAAG TTAGTGGACT GTTTAAGAGA TTTGGGTCCT ACTGATTGGC
1851 AGCTGGCCTG CTTGGTTTGT AAAACTTTAT GGAACTTCAG TGAAAACATC
1901 ACTAATGCTT CGTCATGTTT TGGAAATGAA GACACCAACA CACTCTTACT
1951 CTTGCTCTCA TCATTTTTAG ATGAAGAACT AGCACTGGAT GGCAGTTTTG
2001 ATCCAGACCT AAAAAACTAT CACAAACTCC ATTGGGAAAC AGAATTCAAA
2051 CCTGTGGCAC AGCAGCTTCT AAACCGAATT CAGAGACATC ACACCTTCCT
2101 GGAACCCCTG CCCATTCCCT CTTTCTAACA TGATGCAGAT TAACAGTAGA
2151 AACGAGAACT CACGTCTCCC TCATTCTTAA GAACTGGTAA CAAACGTGAA
2201 CATTTTTTTC AGCATTAACA AATGTGGAAA GTTTTTCAAG AACTGGTTTT
2251 AGTGAGTAGC TGAAGTATTT TTTAAAATTA AGCATTTCTT CTTGTTAGGT
2301 ATTATGGAAA AATGAATATA CACATTATAT TTCCTGTTGA GAGAAATGTA
2351 AGATGAAAAT ATGTGCATTT TCAAGTAAAT GACTTTTTCT TCTATTCTCT
2401 ATTAAACAAT TTAGTTCTAG TCTTAAAAAA AAAAAAAAAA AAAAAAAAAA
2451 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2125 bp; peptide length: 702
Category: putative protein

```
  1 MMGDSMVKIN GIYLTKSNAI CHLKSHPLQL TDDGGFSEIK EQEMFKGTTS
 51 LPSHLKNGGD QGKRHARASS CPSSSDLSRL QTKAVPKADL QEEDAEIEVD
101 EVFWNTRIVP ILRELEKEEN IETVCAACTQ LHHALEEGNM LGNKFKGRSI
151 LLKTLCKLVD VGSDSLSLKL AKIILALKVS RKNLLNVCKL IFKISRNEKN
201 DSLIQNDSIL ESLLEVLRSE DLQTNMEAFL YCMGSIKFIS GNLGFLNEMI
251 SKGAVEILIN LIKQINENIK KCGTFLPNSG HLLVQVTATL RNLVDSSLVR
301 SKFLNISALP QLCTAMEQYK GDKDVCTNIA RIFSKLTSYR DCCTALASYS
351 RCYALFLNLI NKYQKKQDLV VRVVFILGNL TAKNNQAREQ FSKEKGSIQT
401 LLSLFQTFHQ LDLHSQKPVG QRGEQHRAQR PPSEAEDVLI KLTRVLANIA
451 IHPGVGPVLA ANPGIVGLLL TTLEYKSLDD CEELVINATA TINNLSYYQV
501 KNSIIQDKKL YIAELLLKLL VSNNMDGILE AVRVFGNLSQ DHDVCDFIVQ
551 NNVHRFMMAL LDAQHQDICF SACGVLLNLT VDKDKRVILK EGGGIKKLVD
601 CLRDLGPTDW QLACLVCKTL WNFSENITNA SSCFGNEDTN TLLLLLSSFL
651 DEELALDGSF DPDLKNYHKL HWETEFKPVA QQLLNRIQRH HTFLEPLPIP
701 SF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_14p7, frame 2

TREMBL:MMD367_1 product: "KAP3B"; Mus musculus mRNA for KAP3B,
complete cds., N = 2, Score = 97, P = 0.00039

```
>TREMBL:MMD367_1 product: "KAP3B"; Mus musculus mRNA for KAP3B, complete
     cds.
          Length = 772

  HSPs:

 Score = 97 (14.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04
 Identities = 45/163 (27%), Positives = 77/163 (47%)

Query:    442 LTRVLANIAIHPGVGPVLAANPGIVGLLLTTLEYKSLDDCEELVINATATINNLSYYQVK 501
              L +++ NI+ H G  P        VG L   +   S D+ EE VI    T+ NL+   +
Sbjct:    483 LMKMIRNISQHDG--PTKNLFIDYVGDLAAQI---SSDEEEEFVIECLGTLANLTIPDLD 537

Query:    502 -NSIIQDKKLYIAELLLKLLVSNNMDG-ILEAVRVFGNLSQDHDVCDFIVQNNVHRFMMA 559
                 ++++ KL +  L  KL      D  +LE V + G +S D      + ++ +   ++
Sbjct:    538 WELVLKEYKL-VPFLKDKLKPGAAEDDLVLEVVIMIGTVSMDDSCAALLAKSGIIPALIE 596

Query:    560 LLDAQHQDICFSACGVLL---NLTVDKDKR-VILKEGGGIKKLVDCLRD 604
              LL+AQ +D  F  C ++     +   +  R VI+KE      L+D + D
Sbjct:    597 LLNAQQEDDEF-VCQIIYVFYQMVFHQATRDVIIKETQAPAYLIDLMHD 644

 Score = 77 (11:6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04
 Identities = 42/178 (23%), Positives = 82/178 (46%)

Query:    169 KLAKIILALKVSRKNLLNVCK-LIFKISRNEKNDSLIQNDSILESLLEVLRSEDLQTNME 227
              K  K    L V ++ LL V   L+  ++ + + +  ++N +I+  L++ L  +    N E
Sbjct:    263 KTFKKYQGLVVKQEQLLRVALYLLLNLAEDTRTELKMRNKNIVHMLVKALDRD----NFE 318

Query:    228 AFLYCMGSIKFISGNLGFLNEMISKGAVEILINLIKQINENIKKCGTFLPNSGHLLVQVT 287
                +  +  +K +S  +   N+M+    VE L+ +I   +E++            L + +
Sbjct:    319 LLILVVSFLKKLSIFMENKNDMVEMDIVEKLVKMIPCEHEDL------------LNITLR 366

Query:    288 ATLRNLVDSSLVRSKFLNISALPQLCTAM--EQYKGDKDVCT--NIARI--FSKLTSYRD 341
                L    D+ L R+K + +   LP+L  +  E YK    +C   +I+    F  + +Y D
Sbjct:    367 LLLNLSFDTGL-RNKMVQVGLLPKLTALLGNENYK-QIAMCVLYHISMDDRFKSMFAYTD 424
```

```
Query:   342 CCTAL 346
             C   L
Sbjct:   425 CIPQL 429

 Score = 69 (10.4 bits), Expect = 2.6e+00, Sum P(2) = 9.2e-01
 Identities = 35/146 (23%), Positives = 70/146 (47%)

Query:   512 IAELLLKLLVSNNMDGILEAVRVFGNLSQDHDVCDFIVQNNVHRFMMALLDAQHQDICFS 571
             I  +L+K L  +N + ++  V      LS   +  + +V+ ++    ++ ++  +H+D+
Sbjct:   304 IVHMLVKALDRDNFELLILVVSFLKKLSIFMENKNDMVEMDIVEKLVKMIPCEHEDLLNI 363

Query:   572 ACGVLLNLTVDKDKRVILKEGGGIKKLVDCLRDLGPTDW-QLACLVCKTLWNFSENITNA 630
               +LLNL+ D    R  + + G + KL    L   G  ++ Q+A  +C  L++ S +
Sbjct:   364 TLRLLLNLSFDTGLRNKMVQVGLLPKLTALL---GNENYKQIA--MC-VLYHISMD-DRF 416

Query:   631 SSCFGNEDT-NTLLLLLSSFLDEELALD 657
              S F   D    L+ +L     DE + L+
Sbjct:   417 KSMFAYTDCIPQLMKMLFECSDERIDLE 444

 Score = 68 (10.2 bits), Expect = 3.2e-03, Sum P(2) = 3.2e-03
 Identities = 18/58 (31%), Positives = 30/58 (51%)

Query:   190 LIFKISRNEKN-DSLIQNDSILESLLEVLRSE-----DLQTNMEAFLYCMGSIKFISG 241
             LI +++RN  N + L+ N++ L +L  VLR +      +L TN+    +C  S     G
Sbjct:   155 LILQLARNPDNLEELLLNETALGALARVLREDWKQSVELATNIIYIFFCFSSFSHFHG 212

 Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00
 Identities = 26/122 (21%), Positives = 53/122 (43%)

Query:   283 LVQVTATLRNL----VDSSLVRSKFLNISALPQLCTAMEQYKGDKDVCTNIARIFSKLTS 338
             +++    TL NL     +D  LV  ++      +P L    ++    + D+   + I    S
Sbjct:   521 VIECLGTLANLTIPDLDWELVLKEY---KLVPFLKDKLKPGAAEDDLVLEVV-IMIGTVS 576

Query:   339 YRDCCTALASYSRCYALFLNLINKYQKKQDLVVRVVFILGNLTAKNNQAREQFSKEKGSI 398
               D C AL + S        + L+N  Q+  + V +++++   +   +    R+   KE  +
Sbjct:   577 MDDSCAALLAKSGIIPALIELLNAQQEDDEFVCQIIYVFYQMVF-HQATRDVIIKETQAP 635

Query:   399 QTLLSL 404
               L+ L
Sbjct:   636 AYLIDL 641

 Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00
 Identities = 44/177 (24%), Positives = 79/177 (44%)

Query:   481 CE-ELVINATATIN-NLSYYQ-VKNSIIQDKKLYIAELLLKLLVSNNMDGILEAVRVFGN 537
             CE E ++N T  +  NLS+   ++N ++Q    + +   L  LL + N   I  A+ V  +
Sbjct:   355 CEHEDLLNITLRLLLNLSFDTGLRNKMVQ---VGLLPKLTALLGNENYKQI--AMCVLYH 409

Query:   538 LSQDHDVCD-FIVQNNVHRFMMALLDAQHQDICFSACGVLLNLTVDKDKRVILKEGGGIK 596
             +S D       F   + + + M  L +    + I         +NL  +K   ++ EG G+K
Sbjct:   410 ISMDDRFKSMFAYTDCIPQLMKMLFECSDERIDLELISFCINLAANKRNVQLICEGNGLK 469

Query:   597 KLVDCLRDLGPTDWQLACLVCKTLWNFSENITNASSCFGNEDTNTLLLLLSSFLDEELAL 656
             L+   R L   D     L+ K + N S++    + F  +   L   +SS  +EE  +
Sbjct:   470 MLMK--RALKLKD----PLLMKMIRNISQHDGPTKNLF-IDYVGDLAAQISSDEEEEFVI 522

Query:   657 D 657
             +
Sbjct:   523 E 523

 Score = 61 (9.2 bits), Expect = 1.6e-02, Sum P(2) = 1.6e-02
 Identities = 20/66 (30%), Positives = 34/66 (51%)

Query:   304 LNISALPQLCTAM-EQYKGDKDVCTNIARIFSKLTSYRDCCTALASYSRCYALFLNLINK 362
             LN +AL  L   + E +K    ++ TNI  IF   +S+        + Y +  AL +N+I+
Sbjct:   171 LNETALGALARVLREDWKQSVELATNIIYIFFCFSSFSHFHGLITHY-KIGALCMNIIDH 229

Query:   363 YQKKQDL 369
               K+ +L
Sbjct:   230 ELKRHEL 236
```

Pedant information for DKFZphtes3_14p7, frame 2

Report for DKFZphtes3_14p7.2

[LENGTH]	708
[MW]	79266.35
[pI]	6.57

```
[FUNCAT]      30.25 vacuolar and lysosomal organization    [S. cerevisiae, YEL013w] 3e-04
[FUNCAT]      06.04 protein targeting, sorting and translocation   [S. cerevisiae, YEL013w]
3e-04
[FUNCAT]      09.25 vacuolar and lysosomal biogenesis      [S. cerevisiae, YEL013w] 3e-04
[BLOCKS]      BL00923F Aspartate and glutamate racemases proteins
[BLOCKS]      BL00288B Tissue inhibitors of metalloproteinases proteins
[PROSITE]     MYRISTYL          9
[PROSITE]     AMIDATION         1
[PROSITE]     CK2_PHOSPHO_SITE       12
[PROSITE]     PKC_PHOSPHO_SITE       7
[PROSITE]     ASN_GLYCOSYLATION      11
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY     7.49 %
```

```
SEQ   ESKETVMMGDSMVKINGIYLTKSNAICHLKSHPLQLTDDGGFSEIKEQEMFKGTTSLPSH
SEG   ............................................................
PRD   cccceeeeccccceeecccccccccceeeeecccccccccccccchhhhhhhccccccccc

SEQ   LKNGGDQGKRHARASSCPSSSDLSRLQTKAVPKADLQEEDAEIEVDEVFWNTRIVPILRE
SEG   ..............xxxxxxxxxx....................................
PRD   cccccccchhhhhhccccccccchhhhhhhccccchhhhhhhhhhhcccccceehhhhhh

SEQ   LEKEENIETVCAACTQLHHALEEGNMLGNKFKGRSILLKTLCKLVDVGSDSLSLKLAKII
SEG   ..................................................xxxxxxxxxx
PRD   hhhhhcchhhhhhhhhhhhhhhhhcccccccccccccchhhhhheeeeccccchhhhhhh

SEQ   LALKVSRKNLLNVCKLIFKISRNEKNDSLIQNDSILESLLEVLRSEDLQTNMEAFLYCMG
SEG   xxxx........................................................
PRD   hhhhhhhhhhhhhhhhhccccccccccccccccchhhhhhhhhhhcchhhhhhhhhhhcc

SEQ   SIKFISGNLGFLNEMISKGAVEILINLIKQINENIKKCGTFLPNSGHLLVQVTATLRNLV
SEG   ............................................................
PRD   ceeeecccchhhhhhcchhhhhhhhhhhhhhccccccccccccccceeeeeehhhhhhh

SEQ   DSSLVRSKFLNISALPQLCTAMEQYKGDKDVCTNIARIFSKLTSYRDCCTALASYSRCYA
SEG   ............................................................
PRD   ccchhhheeeeccchhhhhhhhhhcccccceeeehhhhhhhhhhcccchhhhhhhhhhhhh

SEQ   LFLNLINKYQKKQDLVVRVVFILGNLTAKNNQAREQFSKEKGSIQTLLSLFQTFHQLDLH
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhheeeeeeecccccchhhhhhhhhhchhhhhhhhhhhhhhhhhhcc

SEQ   SQKPVGQRGEQHRAQRPPSEAEDVLIKLTRVLANIAIHPGVGPVLAANPGIVGLLLTTLE
SEG   ............................................................
PRD   cccccccccccccccccccchhhhhhhhhhhhhhhhccccccceeeccccchhhhhhhhh

SEQ   YKSLDDCEELVINATATINNLSYYQVKNSIIQDKKLYIAELLLKLLVSNNMDGILEAVRV
SEG   .................................xxxxxxxxxxxxx..............
PRD   hhcccchhhhhhhheeeecccccccccceeeehhhhhhhhhhhhhhhhccccchhhhhhhh

SEQ   FGNLSQDHDVCDFIVQNNVHRFMMALLDAQHQDICFSACGVLLNLTVDKDKRVILKEGGG
SEG   ............................................................
PRD   ccccccccceeeeeeecchhhhhhhhhhcccceeeeecceeeeeeeecccceeeeecccc

SEQ   IKKLVDCLRDLGPTDWQLACLVCKTLWNFSENITNASSCFGNEDTNTLLLLLSSFLDEEL
SEG   ...............................................xxxxxxxxxxxxx
PRD   hhhhhhhhcccccccccchhhhhhhcccccccccccccccccccccceeeehhhhhhhhh

SEQ   ALDGSFDPDLKNYHKLHWETEFKPVAQQLLNRIQRHHTFLEPLPIPSF
SEG   xxx.............................................
PRD   hhcccccccccchhhhhhhhhchhhhhhhhhhhhhhhheeeecccccc
```

Prosite for DKFZphtes3_14p7.2

PS00001	206->210	ASN_GLYCOSYLATION	PDOC00001
PS00001	212->216	ASN_GLYCOSYLATION	PDOC00001
PS00001	311->315	ASN_GLYCOSYLATION	PDOC00001
PS00001	385->389	ASN_GLYCOSYLATION	PDOC00001
PS00001	493->497	ASN_GLYCOSYLATION	PDOC00001
PS00001	500->504	ASN_GLYCOSYLATION	PDOC00001
PS00001	543->547	ASN_GLYCOSYLATION	PDOC00001
PS00001	584->588	ASN_GLYCOSYLATION	PDOC00001
PS00001	628->632	ASN_GLYCOSYLATION	PDOC00001
PS00001	632->636	ASN_GLYCOSYLATION	PDOC00001
PS00001	635->639	ASN_GLYCOSYLATION	PDOC00001
PS00005	173->176	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005

PS00005	295->298	PKC_PHOSPHO_SITE	PDOC00005
PS00005	344->347	PKC_PHOSPHO_SITE	PDOC00005
PS00005	387->390	PKC_PHOSPHO_SITE	PDOC00005
PS00005	421->424	PKC_PHOSPHO_SITE	PDOC00005
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	218->222	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	654->658	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00008	17->23	MYRISTYL	PDOC00008
PS00008	64->70	MYRISTYL	PDOC00008
PS00008	144->150	MYRISTYL	PDOC00008
PS00008	384->390	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	580->586	MYRISTYL	PDOC00008
PS00008	641->647	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_14p7.2)

DKFZphtes3_15a13
----

group: testes derived

DKFZphtes3_15a13 encodes a novel 387 amino acid protein with weak similarity to S.cerevisiae Hop1.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to S.cerevisiae Hop1

complete cDNA, complete cds, potential start codon at Bp 116, 3 EST hits
S.cerevisiae Hop1p is a meiosis-specific protein

Sequenced by GBF

Locus: unknown

Insert length: 1848 bp
Poly A stretch at pos. 1766, no polyadenylation signal found

```
   1 GGAAAGCGCA TGCGCGTCGG GCACAGCGCG TGCAGCCTCG TGCAGCTCTT
  51 CTGGTCTCCG GCGCCCGCCC CTCAGACGTA ATGTTGAATT AAAGAAAATA
 101 CTTTATCAGA AGAAGATGGC CACTGCCCAG TTGCAGAGGA CTCCCATGAG
 151 TGCACTGGTA TTTCCCAATA AGATATCAAC TGAACACCAG TCTTTGGTGT
 201 TAGTGAAGAG GCTTCTAGCA GTTTCAGTAT CCTGTATCAC GTATTTGAGG
 251 GGAATATTCC CAGAATGCGC TTATGGAACA AGATATCTAG ATGATCTTTG
 301 TGTCAAAATA CTGAGAGAAG ATAAAAATTG CCCAGGATCT ACACAGTTAG
 351 TGAAATGGAT GCTAGGATGT TATGATGCTT TACAGAAAAA ATATGTATAC
 401 ACAAACCCAG AAGATCCTCA GACAATTTCA GAATGTTACC AATTCAAATT
 451 CAAATACACC AATAATGGAC CACTCATGGA CTTCATAAGT AAAAACCAAA
 501 GCAACGAATC TAGCATGTTG TCTACTGACA CCAAGAAAGC AAGCATTCTC
 551 CTCATTCGCA AGATTTATAT CCTAATGCAA AATCTGGGGC CTTTACCTAA
 601 TGATGTTTGT TTGACCATGA AACTTTTTTA CTATGATGAA GTTACACCCC
 651 CAGATTACCA GCCTCCCGGT TTTAAGGATG GTGATTGTGA AGGAGTTATA
 701 TTTGAAGGGG AACCTATGTA TTTAAATGTG GGAGAAGTCT CAACACCTTT
 751 TCACATCTTC AAAGTAAAAG TGACCACTGA GAGAGAACGA ATGGAAAATA
 801 TTGACTCAAC TATACTATCA CCAAAACAAA TAAAAACACC ATTTCAAAAA
 851 ATCCTGAGGG ACAAAGATGT AGAAGATGAA CAGGAGCATT ATACAAGTGA
 901 TGATTTGGAC ATTGAAACTA AAATGGAAGA ACAGGAAAAA AACCCTGCAT
 951 CTTCTGAACT TGAAGAACCA AGTTTAGTTT GTGAGGAAGA TGAAATTATG
1001 AGGTCTAAAG AAAGTCCAGA TCTTTCTATT TCTCATTCTC AGGTTGAGCA
1051 GTTAGTCAAT AAAACATCTG AACTTGATAT GTCTGAAAGC AAAACAAGAA
1101 GTGGAAAAGT CTTTCAGAAT AAAATGGCAA ATGGAAATCA ACCAGTAAAA
1151 TCTTCCAAAG AAAATCGGAA GAGAAGTCAA CATGAATCTG GGAGAATAGT
1201 CCTCCATCAC TTTGATTCTT CTAGTCAAGA GTCAGTGCCA AAAAGGAGAA
1251 AGTTTAGTGA ACCAAAGGAA CATATATAAA AATTATTTTT GTTCTGCAGG
1301 CTTGCAGAGT TCTTCTCACC ATTTAAACTG AAGGACCCTA TATTATATTT
1351 CCCTAACTCT GAAGATGTAT ATGTAGTTTA AAGCAGTTTG TACACTAAAA
1401 CTAAGTTTTT GGCTGACTGT CATATTGTGG TCCTTAATCT TGAGATAAAT
1451 CCAATAGAAC TTTTGAATAA AAGCAAAAGT ACAAATGTCA TAATTGATTC
1501 GGTAATAAGT AAAATTTCAA AATTGATTTT GTTCATTACC TACTTAATAT
1551 TTCCTTTAAA TATATACTAA CTGTTAAGGC CCTCTAATGC CATTTTTCTA
1601 AACAGTAATG TTTACTTTGG TATTAAAATT TGGTATGGAT TCACTTTTTA
1651 CTTATGTTAA AATTATACCA TTTAACTGGC TCTTTTGTCA TTGTGCTGTT
1701 ATTAAAACAA TGTTCTTCAA TATTTTGACA TAATGTATTA ACATTTTAAT
1751 ATATAATGTA CAATTTAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAGG
1801 GGCGGCCGCT CTAGAGGATC CAAGCTTACG TACAAAAAAA AAAAAAGG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 116 bp to 1276 bp; peptide length: 387
Category: similarity to known protein

```
  1 MATAQLQRTP MSALVFPNKI STEHQSLVLV KRLLAVSVSC ITYLRGIFPE
 51 CAYGTRYLDD LCVKILREDK NCPGSTQLVK WMLGCYDALQ KKYVYTNPED
101 PQTISECYQF KFKYTNNGPL MDFISKNQSN ESSMLSTDTK KASILLIRKI
151 YILMQNLGPL PNDVCLTMKL FYYDEVTPPD YQPPGFKDGD CEGVIFEGEP
201 MYLNVGEVST PFHIFKVKVT TERERMENID STILSPKQIK TPFQKILRDK
251 DVEDEQEHYT SDDLDIETKM EEQEKNPASS ELEEPSLVCE EDEIMRSKES
301 PDLSISHSQV EQLVNKTSEL DMSESKTRSG KVFQNKMANG NQPVKSSKEN
351 RKRSQHESGR IVLHHFDSSS QESVPKRRKF SEPKEHI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15a13, frame 2

TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence., N = 1, Score = 274, P = 5.7e-22

TREMBL:SC9877_9 gene: "hop1"; S.cerevisiae chromosome IX cosmid 9877., N = 2, Score = 126, P = 7.1e-09

PIR:A34691 meiosis-specific protein HOP1 - yeast (Saccharomyces cerevisiae), N = 2, Score = 126, P = 7.8e-08

```
>TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from
     Arabidopsis thaliana chromosome 1, complete sequence.
            Length = 562

  HSPs:

 Score = 274 (41.1 bits), Expect = 5.7e-22, P = 5.7e-22
 Identities = 84/290 (28%), Positives = 145/290 (50%)

Query:    22 TEHQSLVLVKRLLAVSVSCITYLRGIFPECAYGTRYLDDLCVKILREDKNCPGSTQLVKW 81
             TE  SL+L + LL +++  I+Y+RG+FPE  +  + +  L +KI +        S +L+ W
Sbjct:    11 TEQDSLLLTRNLLRIAIFNISYIRGLFPEKYFNDKSVPALDMKIKKLMPMDAESRRLIDW 70

Query:    82 M-LGCYDALQKKYVYT------NPEDPQTISECYQFKFKYTNNGP--LMDFISK--NQSN 130
             M  G YDALQ+KY+ T        D   I E Y F F Y+++   +M  I++  N+ N
Sbjct:    71 MEKGVYDALQRKYLKTLMFSICETVDGPMIEE-YSFSFSYSDSDSQDVMMNINRTGNKKN 129

Query:   131 ESSMLST------DTKKASILLIRKIYILMQNLGPLPNDVCLTMKLFYYDEVTPPDYQPP 184
                  ST        + ++  ++R +  LM+ L  +P++  + MKL YYD+VTPPDY+PP
Sbjct:   130 GGIFNSTADITPNQMRSSACKMVRTLVQLMRTLDKMPDERTIVMKLLYYDDVTPPDYEPP 189

Query:   185 GFKD--GDCEGVIFEGEPMYLNVGEVSTPFHIFKVKVTT-------ERERMENIDSTILS 235
             F+     D     ++    P+ + +G V++    +  +KV +         E + M++ D   +
Sbjct:   190 FFRGCTEDEAQYVWTKNPLRMEIGNVNSKHLVLTLKVKSVLDPCEDENDDMQD-DGKSIG 248

Query:   236 PKQIKTPFQKILRDKDVEDEQEHY-------TSDDLDIETKMEEQEKNPASSE 281
             P  +     Q   D ++   QE+           DD D E    ++   ++PA +E
Sbjct:   249 PDSVHDD-QPSDSDSEISQTQENQFIVAPVEKQDDDDGEVDEDDNTQDPAENE 300
```

Pedant information for DKFZphtes3_15a13, frame 2

Report for DKFZphtes3_15a13.2

```
[LENGTH]      387
[MW]          44417.64
[pI]          5.57
[HOMOL]       TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from
Arabidopsis thaliana chromosome 1, complete sequence. 9e-23
[FUNCAT]      09.13 biogenesis of chromosome structure    [S. cerevisiae, YIL072w] 7e-11
[FUNCAT]      03.19 recombination and dna repair    [S. cerevisiae, YIL072w] 7e-11
[FUNCAT]      03.13 meiosis  [S. cerevisiae, YIL072w] 7e-11
[FUNCAT]      30.10 nuclear organization    [S. cerevisiae, YIL072w] 7e-11
[PIRKW]       nucleus 2e-09
[PIRKW]       zinc finger 2e-09
```

[PIRKW]	DNA binding 2e-09	
[PROSITE]	MYRISTYL	1
[PROSITE]	CAMP_PHOSPHO_SITE	3
[PROSITE]	CK2_PHOSPHO_SITE	12
[PROSITE]	PKC_PHOSPHO_SITE	7
[PROSITE]	ASN_GLYCOSYLATION	3
[KW]	Alpha_Beta	

```
SEQ   MATAQLQRTPMSALVFPNKISTEHQSLVLVKRLLAVSVSCITYLRGIFPECAYGTRYLDD
PRD   cccccccccccccccccccchhhhhhhhhhhhhhhhhhhheeeeeccccccccccccchh

SEQ   LCVKILREDKNCPGSTQLVKWMLGCYDALQKKYVYTNPEDPQTISECYQFKFKYTNNGPL
PRD   hhhhhhcccccccccccccccchhhhhhhhhhhcccccccchhhhheeeeeeccccccce

SEQ   MDFISKNQSNESSMLSTDTKKASILLIRKIYILMQNLGPLPNDVCLTMKLFYYDEVTPPD
PRD   eeeecccccccceeeecccchhhhhhhhhhhhhhhcccccccccceeeeeeeeeeccccc

SEQ   YQPPGFKDGDCEGVIFEGEPMYLNVGEVSTPFHIFKVKVTTERERMENIDSTILSPKQIK
PRD   ccccccccccccceeeeeccceeeeecccccccceeeeeeccccchhhhcccccccccchhh

SEQ   TPFQKILRDKDVEDEQEHYTSDDLDIETKMEEQEKNPASSELEEPSLVCEEDEIMRSKES
PRD   hhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhccccccccccccchhhhhhhhhhcc

SEQ   PDLSISHSQVEQLVNKTSELDMSESKTRSGKVFQNKMANGNQPVKSSKENRKRSQHESGR
PRD   ccccccchhhhhhhhhccccccccccccccceeeeecccccccccchhhhhhhhhcccce

SEQ   IVLHHFDSSSQESVPKRRKFSEPKEHI
PRD   eeeeccccccccccccccccccccccc
```

Prosite for DKFZphtes3_15a13.2

PS00001	127->131	ASN_GLYCOSYLATION	PDOC00001
PS00001	130->134	ASN_GLYCOSYLATION	PDOC00001
PS00001	315->319	ASN_GLYCOSYLATION	PDOC00001
PS00004	140->144	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	139->142	PKC_PHOSPHO_SITE	PDOC00005
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00005	221->224	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00005	329->332	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00006	96->100	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	177->181	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	260->264	CK2_PHOSPHO_SITE	PDOC00006
PS00006	268->272	CK2_PHOSPHO_SITE	PDOC00006
PS00006	280->284	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00006	346->350	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00008	84->90	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_15a13.2)

DKFZphtes3_15c24

group: metabolism

DKFZphtes3_15c24 encodes a novel 404 amino acid protein with strong similarity to 2-hydroxyacid dehydrogenases.

The novel protein contains a D-isomer specific 2-hydroxyacid dehydrogenases signature. Proteins with such a signature have similar enzymatic activities: D-lactate dehydrogenase (EC 1.1.1.28), catalyzes the reduction of D-lactate to pyruvate. D-glycerate dehydrogenase (EC 1.1.1.29) catalyzes the reduction of hydroxypyruvate to glycerate. 3-phosphoglycerate dehydrogenase (EC 1.1.1.95), catalyzes the oxidation of D-3-phosphoglycerate to 3-phosphohydroxypyruvate. Therefore the novel protein is a new 2-hydroxyacid dehydrogenase.

The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

strong similarity to C.elegans T03F1.1

potential start at Bp 55 matches kozak consensus PyCCatgG

Sequenced by GBF

Locus: unknown

Insert length: 1956 bp
Poly A stretch at pos. 1929, polyadenylation signal at pos. 1903

```
   1 CGAAGGCGGC GGCGAAGGCC CGGGCTGGGA GCGTTGGCGG CCGGAGTCCC
  51 AGCCATGGCG GAGTCTGTGG AGCGCCTGCA GCAGCGGGTC CAGGAGCTGG
 101 AGCGGGAACT TGCCCAGGAG AGGAGTCTGC AGGTCCCGAG GAGCGGCGAC
 151 GGAGGGGGCG GCCGGGTCCG CATCGAGAAG ATGAGCTCAG AGGTGGTGGA
 201 TTCGAATCCC TACAGCCGCT TGATGGCATT GAAACGAATG GGAATTGTAA
 251 GCGACTATGA GAAAATCCGT ACCTTTGCCG TAGCAATAGT AGGTGTTGGT
 301 GGAGTAGGTA GTGTGACTGC TGAAATGCTG ACAAGATGTG GCATTGGTAA
 351 GTTGCTACTC TTTGATTATG ACAAGGTGGA ACTAGCCAAT ATGAATAGAC
 401 TTTTCTTCCA ACCTCATCAA GCAGGATTAA GTAAAGTTCA AGCAGCAGAA
 451 CATACTCTGA GGAACATTAA TCCTGATGTT CTTTTTGAAG TACACAACTA
 501 TAATATAACC ACAGTGGAAA ACTTTCAACA TTTCATGGAT AGAATAAGTA
 551 ATGGTGGGTT AGAAGAAGGA AAACCTGTTG ATCTAGTTCT TAGCTGTGTG
 601 GACAATTTTG AAGCTCGAAT GACAATAAAT ACAGCTTGTA ATGAACTTGG
 651 ACAAACATGG ATGGAATCTG GGGTCAGTGA AAATGCAGTT TCAGGGCATA
 701 TACAGCTTAT AATTCCTGGA GAATCTGCTT GTTTTGCGTG TGCTCCACCA
 751 CTTGTAGTTG CTGCAAATAT TGATGAAAAG ACTCTGAAAC GAGAAGGTGT
 801 TTGTGCAGCC AGTCTTCCTA CCACTATGGG TGTGGTTGCT GGGATCTTAG
 851 TACAAAACGT GTTAAAGTTT CTGTTAAATT TTGGTACTGT TAGTTTTTAC
 901 CTTGGATACA ATGCAATGCA GGATTTTTTT CCTACTATGT CCATGAAGCC
 951 AAATCCTCAG TGTGATGACA GAAATTGCAG GAAGCAGCAG GAGGAATATA
1001 AGAAAAAGGT AGCAGCACTG CCTAAACAAG AGGTTATACA AGAAGAGGAA
1051 GAGATAATCC ATGAAGATAA TGAATGGGGT ATTGAGCTGG TATCTGAGGT
1101 TTCAGAAGAG GAACTGAAAA ATTTTTCAGG TCCAGTTCCA GACTTACCTG
1151 AAGGAATTAC AGTGGCATAC ACAATTCCAA AAAAGCAAGA AGATTCTGTC
1201 ACTGAGTTAA CAGTGGAAGA TTCTGGTGAA AGCTTGGAAG ACCTCATGGC
1251 CAAAATGAAG AATATGTAGA TAATGGACTG GGATATATTG TATTTCTCAT
1301 GTTAAAGCCT CTTCCCTTGA AATTAAAAAA AAATTTTAAC TGATAAAACT
1351 TAGGGCAACA TTAATTAATG TATATTCTTA CCTGAATTGT TATACTTTTT
1401 GAAAATCCTG TGACTTGCCT GTTTCTCCCC GCTCCAACGA AATCATTAAC
1451 TCTCCTAAAA TGTGTTTCAT TCTAGTAAGA AAACCTCAAA GGATATTGTA
1501 GGATATAAAT CTTACTTGAA AACATAGCTG TTGAAATGTT TTGGCCTTTT
1551 GGAGTGGGGG AAGGACAAAT CTGATCCTGT AATCTTTTTC TTTCCAGTAA
1601 TCCCTTGTGT CTGTTGCATG AGGACATGGA CAATAAAGTA GTATATGATC
1651 CTCAGATACA GGGAGAAGGA CAAGGCATAC AGCTTATTGA TTAGAGCTGG
1701 CAAGCATCTG CTCATTATGT TTGGAATTGC TTTCTATAAG AAAATTGCCC
1751 ACTACTACTA ACTTGATCAA CAATGAATTC AAAATAGTTA ACCTATGAAA
1801 TAACATCCTC TCAAATGTTT GCTGATGAAG TACAAGTTGA AATGTAGTTA
1851 TTGGAAAAGT CTGTAACCTG TGGATCATAT ATATTCAAAG TGAGACAAAG
1901 GCAAATAAAA AGCAGCTATT TTCATGAATA GACAAAAAAA AAAAAAAAAA
1951 AAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 55 bp to 1266 bp; peptide length: 404
Category: similarity to unknown protein
Classification: Metabolism
Prosite motifs: D_2_HYDROXYACID_DH_1 (76-105)

```
  1 MAESVERLQQ RVQELERELA QERSLQVPRS GDGGGGRVRI EKMSSEVVDS
 51 NPYSRLMALK RMGIVSDYEK IRTFAVAIVG VGGVGSVTAE MLTRCGIGKL
101 LLFDYDKVEL ANMNRLFFQP HQAGLSKVQA AEHTLRNINP DVLFEVHNYN
151 ITTVENFQHF MDRISNGGLE EGKPVDLVLS CVDNFEARMT INTACNELGQ
201 TWMESGVSEN AVSGHIQLII PGESACFACA PPLVVAANID EKTLKREGVC
251 AASLPTTMGV VAGILVQNVL KFLLNFGTVS FYLGYNAMQD FFPTMSMKPN
301 PQCDDRNCRK QQEEYKKKVA ALPKQEVIQE EEEIIHEDNE WGIELVSEVS
351 EEELKNFSGP VPDLPEGITV AYTIPKKQED SVTELTVEDS GESLEDLMAK
401 MKNM
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15c24, frame 1

TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid T03F1., N = 1, Score = 1204, P = 1.9e-122

TREMBL:ATAC98_3 gene: "YUP8H12.3"; Arabidopsis thaliana chromosome 1 YAC yUP8H12 complete sequence., N = 1, Score = 733, P = 1.5e-72

PIR:A69319 thiamine biosynthesis protein (thiF) homolog - Archaeoglobus fulgidus, N = 1, Score = 218, P = 1.8e-17

TREMBL:AF022796_4 gene: "moeB"; product: "MoeB"; Staphylococcus carnosus molybdenum cofactor biosynthetic gene cluster, complete sequence., N = 1, Score = 220, P = 3.7e-16

>TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid T03F1.
Length = 419

HSPs:

Score = 1204 (180.6 bits), Expect = 1.9e-122, P = 1.9e-122
Identities = 241/367 (65%), Positives = 293/367 (79%)

```
Query:    37 RVRIEKMSSEVVDSNPYSRLMALKRMGIVSDYEKIRTFAVAIVGVGGVGSVTAEMLTRCG 96
             R +IEK+S+EVVDSNPYSRLMAL+RMGIV++YE+IR   VA+VGVGGVGSV AEMLTRCG
Sbjct:    48 RQKIEKLSAEVVDSNPYSRLMALQRMGIVNEYERIREKTVAVVGVGGVGSVVAEMLTRCG 107

Query:    97 IGKLLLFDYDKVELANMNRLFFQPHQAGLSKVQAAEHTLRNINPDVLFEVHNYNITTVEN 156
             IGKL+LFDYDKVE+ANMNRLF+QP+QAGLSKV+AA  TL ++NPDV  EVHN+NITT++N
Sbjct:   108 IGKLILFDYDKVEIANMNRLFYQPNQAGLSKVEAARDTLIHVNPDVQIEVHNFNITTMDN 167

Query:   157 FQHFMDRISNGGLEEGKPVDLVLSCVDNFEARMTINTACNELGQTWMESGVSENAVSGHI 216
             F  F++RI  G L +GK +DLVLSCVDNFEARM +N ACNE  Q WMESGVSENAVSGHI
Sbjct:   168 FDTFVNRIRKGSLTDGK-IDLVLSCVDNFEARMAVNMACNEENQIWMESGVSENAVSGHI 226

Query:   217 QLIIPGESACFACAPPLVVAANIDEKTLKREGVCAASLPTTMGVVAGILVQNVLKFLLNF 276
             Q I PG++ACFAC PPLVVA+ IDE+TLKR+GVCAASLPTTM VVAG LV N LK+LLNF
Sbjct:   227 QYIEPGKTACFACVPPLVVASGIDERTLKRDGVCAASLPTTMAVVAGFLVMNTLKYLLNF 286

Query:   277 GTVSFYLGYNAMQDFFPTMSMKPNPQCDDRNCRKQQEEYKKKVAALPKQ-EV-IQEEEEI 334
             G VS Y+GYNA+ DFFP  S+KPNP CDD +C ++Q+EY++KVA  P   EV + EEE +
Sbjct:   287 GEVSQYVGYNALSDFFPRDSIKPNPYCDDSHCLQRQKEYEEKVANQPVDLEVEVPEEETV 346

Query:   335 IHEDNEWGIELVSEVSEEELKNFSGPVPDLPEGITVAYTIPKKQEDSVTELTVEDSGESL 394
             +HEDNEWGIELV+E SE     +   S     +     G+  AY   P K+  D+  TEL+     +   +
Sbjct:   347 VHEDNEWGIELVNE-SEPSAEQSSSL--NAGTGLKFAYE-PIKR-DAQTELSPAQA--AT 399

Query:   395 EDLMAKMKN 403
```

```
              D M  +K+
Sbjct:    400 HDFMKSIKD 408
```

Pedant information for DKFZphtes3_15c24, frame 1

Report for DKFZphtes3_15c24.1

```
[LENGTH]        404
[MW]            44863.36
[pI]            4.79
[HOMOL]         TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid T03F1. 1e-115
[FUNCAT]        h cofactor metabolism         [H. influenzae, HI1449] 2e-08
[FUNCAT]        06.07 protein modification (glycolsylation, acylation, myristylation,
palmitylation, farnesylation and processing)        [S. cerevisiae, YDR390c UBA2 - E1-like]
4e-07
[FUNCAT]        04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation)  [S.
cerevisiae, YDR390c UBA2 - E1-like] 4e-07
[FUNCAT]        06.13.01 cytoplasmic degradation      [S. cerevisiae, YDR390c UBA2 - E1-like]
4e-07
[FUNCAT]        30.10 nuclear organization    [S. cerevisiae, YDR390c UBA2 - E1-like] 4e-07
[FUNCAT]        11.01 stress response  [S. cerevisiae, YKL210w UBA1 - E1-like] 2e-06
[FUNCAT]        30.03 organization of cytoplasm       [S. cerevisiae, YKL210w UBA1 - E1-like]
2e-06
[BLOCKS]        BL01042A Homoserine dehydrogenase proteins
[PIRKW]         thiamine pyrophosphate 1e-07
[PIRKW]         molybdenum 5e-07
[PIRKW]         molybdopterin biosynthesis 5e-07
[SUPFAM]        molybdopterin biosynthesis protein moeB 2e-12
[PROSITE]       D_2_HYDROXYACID_DH_1   1
[KW]            TRANSMEMBRANE  1
[KW]            LOW_COMPLEXITY     8.66 %
```

```
SEQ  MAESVERLQQRVQELERELAQERSLQVPRSGDGGGGRVRIEKMSSEVVDSNPYSRLMALK
SEG  ............................................................
PRD  ccchhhhhhhhhhhhhhhhhhhhhhhccccccccccccceeeccccccccccchhhhhhc
MEM  ............................................................

SEQ  RMGIVSDYEKIRTFAVAIVGVGGVGSVTAEMLTRCGIGKLLLFDYDKVELANMNRLFFQP
SEG  ................xxxxxxxxx...................................
PRD  cccccchhhhhheeeeecccccchhhhhhhhhcccceeeeecccccchhhhhhhhhhhhc
MEM  ........MMMMMMMMMMMMMMMMMMMMMMM.............................

SEQ  HQAGLSKVQAAEHTLRNINPDVLFEVHNYNITTVENFQHFMDRISNGGLEEGKPVDLVLS
SEG  ............................................................
PRD  ccccchhhhhhhhhhhhccccceeeeecccccchhhhhhhhhhhccccccccccceeeee
MEM  ............................................................

SEQ  CVDNFEARMTINTACNELGQTWMESGVSENAVSGHIQLIIPGESACFACAPPLVVAANID
SEG  ............................................................
PRD  cccchhhhhhhhhhhhhcccccccccccccccccceeeecccccccceeecccccccccc
MEM  ............................................................

SEQ  EKTLKREGVCAASLPTTMGVVAGILVQNVLKFLLNFGTVSFYLGYNAMQDFFPTMSMKPN
SEG  ............................................................
PRD  ccccccccccccccccchhhhhhhhhhhhhhhhcccceeecccccccccccccccccccc
MEM  ............................................................

SEQ  PQCDDRNCRKQQEEYKKKVAALPKQEVIQEEEEIIHEDNEWGIELVSEVSEEELKNFSGP
SEG  ........................xxxxxxxxxxxxxxx...xxxxxxxxxxx.......
PRD  ccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccceeeeeehhhhhhhhhcccc
MEM  ............................................................

SEQ  VPDLPEGITVAYTIPKKQEDSVTELTVEDSGESLEDLMAKMKNM
SEG  ............................................
PRD  ccccccceeeeeeehhhhhhheeeeecccccchhhhhhhhhccc
MEM  ............................................
```

Prosite for DKFZphtes3_15c24.1

```
PS00065     76->105    D_2_HYDROXYACID_DH_1   PDOC00063
```

(No Pfam data available for DKFZphtes3_15c24.1)

DKFZphtes3_15c6
----

group: transmembrane protein

DKFZphtes3_15c6 encodes a novel 118 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1283 bp
Poly A stretch at pos. 1264, no polyadenylation signal found

```
   1 GAGACACTGA GCCCCGAGAC AGTGAGTGGT GGCCTCACTG CTCTGCCCGG
  51 CACCCTGTCA CCTCCACTTT GCCTTGTTGG AAGTGACCCA GCCCCCTCCC
 101 CTTCCATTCT CCCACCTGTT CCCCAGGACT CACCCCAGCC CCTGCCTGCC
 151 CCTGAGGAAG AAGAGGCACT CACCACTGAG GACTTTGAGT TGCTGGATCA
 201 GGGGGAGCTG GAGCAGCTGA ATGCAGAGCT GGGCTTGGAG CCAGAGACAC
 251 CGCCAAAACC CCCTGATGCT CCACCCCTGG GGCCCGACAT CCATTCTCTG
 301 GTACAGTCAG ACCAAGAAGC TCAGGCCGTG GCAGAGCCAT GAGCCAGCCG
 351 TTGAGGAAGG AGCTGCAGGC ACAGTAGGGC TTCCTGGCTA GGAGTGTTGC
 401 TGTTTCCTCC TTTGCCTACC ACTCTGGGGT GGGGCAGTGT GTGGGGAAGC
 451 TGGCTGTCGG ATGGTAGCTA TTCCACCCTC TGCCTGCCTG CCTGCCTGCT
 501 GTCCTGGGCA TGGTGCAGTA CCTGTGCCTA GGATTGGTTT TAAATTTGTA
 551 AATAATTTTC CATTTGGGTT AGTGGATGTG AACAGGGCTA GGGAAGTCCT
 601 TCCCACAGCC TGCGCTTGCC TCCCTGCCTC ATCTCTATTC TCATTCCACT
 651 ATGCCCCAAG CCCTGGTGGT CTGGCCCTTT CTTTTTCCTC CTATCCTCAG
 701 GGACCTGTGC TGCTCTGCCC TCATGTCCCA CTTGGTTGTT TAGTTGAGGC
 751 ACTTTATAAT TTTTCTCTTG TCTTGTGTTC CTTTCTGCTT TATTTCCCTG
 801 CTGTGTCCTG TCCTTAGCAG CTCAACCCCA TCCTTTGCCA GCTCCTCCTA
 851 TCCCGTGGGC ACTGGCCAAG CTTTAGGGAG GCTCCTGGTC TGGGAAGTAA
 901 AGAGTAAACC TGGGGCAGTG GGTCAGGCCA GTAGTTACAC TCTTAGGTCA
 951 CTGTAGTCTG TGTAACCTTC ACTGCATCCT TGCCCCATTC AGCCCGGCCT
1001 TTCATGATGC AGGAGAGCAG GGATCCCGCA GTACATGGCG CCAGCACTGG
1051 AGTTGGTGAG CATGTGCTCT CTCTTGAGAT TAGGAGCTTC CTTACTGCTC
1101 CTCTGGGTGA TCCAAGTGTA GTGGGACCCC CTACTAGGGT CAGGAAGTGG
1151 ACACTAACAT CTGTGCAGGT GTTGACTTGA AAAATAAAGT GTTGATTGGC
1201 TAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAGGGCGGCC GCTCTAGAGG
1251 ATCCAAGCTT ACGTAAAAAA AAAAAAAAAA AAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 461 bp to 814 bp; peptide length: 118
Category: putative protein

```
  1 MVAIPPSACL PACCPGHGAV PVPRIGFKFV NNFPFGLVDV NRAREVLPTA
 51 CACLPASSLF SFHYAPSPGG LALSFSSYPQ GPVLLCPHVP LGCLVEALYN
101 FSLVLCSFLL YFPAVSCP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15c6, frame 2

PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana, N = 1, Score = 76, P = 0.33

```
>PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana
            Length = 258

  HSPs:

 Score = 76 (11.4 bits), Expect = 4.0e-01, P = 3.3e-01
 Identities = 30/91 (32%), Positives = 44/91 (48%)

Query:    15 PGHGAVPVPRIGFKFVNNFPFGLVDVNRAREVLPTACACLPASSLFSFHYAPSPGGLALS 74
             PG GA P+ R+ F+      PF       + +E+   A  C P SSL+      A    G  L
Sbjct:    52 PGRGA-PLARVTFRH----PFRF---KKQKELFVAAEVCTPVSSLYCGKKATLVVGNVLP 103

Query:    75 FSSYPQGPVLLCP---HV-PLGCLVEALYNFSLVL 105
               S P+G V+ C    HV  G L  A  ++++V+
Sbjct:   104 LRSIPEGAVV-CNVEHHVGDRGVLARASGDYAIVI 137
```

Pedant information for DKFZphtes3_15c6, frame 2

Report for DKFZphtes3_15c6.2

```
[LENGTH]     118
[MW]         12413.79
[pI]         7.53
[PROSITE]    LEUCINE_ZIPPER 1
[PROSITE]    MYRISTYL       1
[PROSITE]    ASN_GLYCOSYLATION      1
[KW]         TRANSMEMBRANE  1

SEQ    MVAIPPSACLPACCPGHGAVPVPRIGFKFVNNFPFGLVDVNRAREVLPTACACLPASSLF
PRD    cccccccccccccccccccccccccceeeecccccceeehhhhhhcccccceeeccccccc
MEM    ............................................................

SEQ    SFHYAPSPGGLALSFSSYPQGPVLLCPHVPLGCLVEALYNFSLVLCSFLLYFPAVSCP
PRD    eeecccccccceeeeecccccccccccccccchhhhhhhcchhhhhhhhccccccccc
MEM    ......................................MMMMMMMMMMMMMMMMMM.
```

Prosite for DKFZphtes3_15c6.2

PS00001	100->104	ASN_GLYCOSYLATION	PDOC00001
PS00008	70->76	MYRISTYL	PDOC00008
PS00029	84->106	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_15c6.2)

DKFZphtes3_15g14
----

group: testes derived

DKFZphtes3_15g14 encodes a novel 701 amino acid protein with weak similarity to S. cerevisiae hypothetical protein YOR243c.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YOR243c

complete cDNA, complete cds, potential start codon at Bp 35, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3495 bp
Poly A stretch at pos. 3462, no polyadenylation signal found

```
   1 GCCTTCCACT GAACCGAGGC ACTGTTATAG AAGAATGGAA GAAGATACAG
  51 ATTATAGAAT CAGGTTTAGT TCTTTGTGTT TCTTTAATGA TCACGTTGGA
 101 TTTCATGGCA CTATAAAAAG CTCACCAAGT GACTTTATTG TTATTGAAAT
 151 TGATGAACAG GGACAGTTAG TTAATAAGAC CATCGATGAG CCTATTTTCA
 201 AGATTAGTGA AATACAACTT GAGCCAAATA ATTTTCCCAA AAAACCAAAA
 251 CTAGATCTTC AAAATCTGTC CTTAGAAGAT GGAAGAAACC AAGAAGTTCA
 301 TACTTTGATT AAGTACACTG ATGGTGACCA AAATCATCAG TCTGGTTCAG
 351 AAAAGGAAGA TACTATCGTT GATGGAACTT CCAAATGTGA AGAAAAAGCT
 401 GATGTTTTAA GCTCCTTTTT GGATGAAAAA ACTCATGAGT TACTGAATAA
 451 TTTTGCCTGT GATGTAAGAG AGAAGTGGCT TTCTAAAACA GAGCTAATTG
 501 GACTACCTCC TGAATTCTCA ATAGGCAGAA TCCTTGACAA AAACCAGAGG
 551 GCTAGTTTAC ACAGTGCCAT TAGGCAGAAA TTTCCATTTT TAGTAACTGT
 601 AGGAAAAAAC AGTGAAATTG TTGTAAAACC AAATCTTGAA TATAAAGAAC
 651 TTTGTCATTT GGTATCTGAA GAGGAAGCAT TTGACTTTTT TAAATATTTG
 701 GATGCAAAGA AAGAAAATTC CAAATTTACC TTTAAACCTG ATACAAACAA
 751 AGACCACAGA AAAGCTGTCC ACCATTTTGT CAACAAAAAG TTTGGAAACC
 801 TTGTGGAAAC CAAATCTTTT TCTAAAATGA ATTGCAGTGC TGGTAATCCG
 851 AATGTGGTGG TAACAGTAAG ATTTCGGGAA AAAGCACACA AACGTGGGAA
 901 AAGGCCTCTT TCTGAATGCC AAGAAGGAAA AGTTATATAT ACAGCTTTTA
 951 CCCTACGAAA GGAAAACCTG GAAATGTTTG AAGCGATTGG TTTTTTAGCT
1001 ATCAAACTTG GTGTTATTCC TTCGGATTTT AGTTATGCAG GCCTTAAAGA
1051 CAAGAAAGCC ATCACCTATC AAGCAATGGT TGTTAGAAAA GTGACTCCAG
1101 AGAGGTTGAA AAATATTGAA AAAGAAATTG AAAAGAAAAG AATGAATGTC
1151 TTTAATATTC GGTCTGTAGA TGATTCCCTG AGACTTGGTC AGCTCAAAGG
1201 AAATCACTTT GATATTGTCA TTAGAAATTT AAAAAAACAA ATAAATGATT
1251 CTGCAAACCT GAGGGAGAGA ATTATGGAAG CAATAGAAAA TGTTAAGAAA
1301 AAAGGCTTTG TGAATTACTA TGGACCACAG AGATTTGGGA AGGGAAGGAA
1351 AGTTCACACA GACCAAATTG GACTAGCTTT GCTGAAGAAT GAAATGATGA
1401 AAGCCATAAA ATTGTTTCTT ACACCAGAAG ACTTGGATGA TCCTGTAAAT
1451 AGAGCAAAGA AGTATTTTCT TCAAACTGAG GATGCTAAAG GCACACTTTC
1501 ATTGATGCCT GAATTCAAAG TGCGTGAGAG AGCATTGTTG GAGGCATTGC
1551 ACCGCTTTGG CATGACCGAG GAAGGTTGTA TCCAGGCATG GTTCTCTTTA
1601 CCCCATTCCA TGCGCATATT CTATGTTCAC GCATATACCA GCAAAATTTG
1651 GAATGAGGCA GTATCTTACA GACTTGAAAC CTATGGAGCA AGAGTAGTGC
1701 AGGGTGATTT GGTCTGTTTG GATGAAGACA TTGATGACGA GAATTTCCCA
1751 AATAGTAAAA TTCACCTGGT AACTGAAGAG GAGGGATCAG CTAATATGTA
1801 TGCAATACAT CAGGTGGTTC TTCCAGTACT TGGATACAAT ATTCAGTACC
1851 CGAAGAACAA AGTAGGGCAG TGGTACCATG ACATACTTAG CAGAGATGGA
1901 CTACAGACAT GTAGGTTTAA AGTACCTACT CTGAAACTGA ATATACCAGG
1951 TTGCTATAGA CAGATTTTGA AACATCCCTG TAATCTCTCA TACCAACTAA
2001 TGGAAGATCA TGACATTGAT GTCAAAACGA AAGGTTCCCA CATTGATGAA
2051 ACAGCTTTGT CTCTTTTGAT CTCTTTTGAT CTTGATGCTT CATGCTATGC
2101 TACCGTTTGT CTGAAGGAAA TAATGAAGCA TGACGTTTAA AACTGATACC
2151 CTTGGTATAA CCATATATAT GTCACCCTTT CCTGTTTTTG AAATTATTGA
2201 TCAGAACAAT ATACAAGGGA AATGCCATAC CTCTGTTTGT GATAGATACC
2251 CCAGAGTAGT TATTACCTCT TTGTGAGATA AGTAATCTTT GATGAAGATT
2301 GAAATACAAT TTCTCATCCA ATTTTTATAT CTTGGCATAC GCTGACCCTC
2351 TTGACCATTT GTAATTTTTT CATATTATCT AAAACAGGTG TTAGAGTCAG
2401 ACAGATTCAT TCTTAGATTC TAGCTCTGAC ACTTACTAGT GATTTTGAGT
2451 ATGTTGTTGA TTTTTTTGTG TGTGGTTACT GATAGAATCA AGACAATTAC
2501 AACTTCATAA ATGACAAATA ATAGGATTAT CTCCACATTT TCTGTTGCTG
2551 GAGGAACAAA ACATTGTGCC CATTTGAAAA TTTTAATTTT TGTTGGTTTA
2601 ACTATCCCAC ATTATAAATC ATCCTTCACC ATTTTATATC AGTTAAATAT
2651 GGGTGTGTTG GGGAGGAATG ACTGGCATGT AGACATGTAT TGATTTAGGA
2701 AGATCTGAGC ATTTCTTTCA TTGTTGGTAA GATATAATGA TGAAATTTAA
```

```
2751 AAAGCAGTAT GGAGCATTAT ATATCAGTAA TGTGATATAT ATACTTAAGC
2801 CAGTTTAACC ATTTTGGGAA ATGTTAGCAT TAGGAAATAA AATCCAAAAG
2851 AAGGAAGAGA AGCTATATGC AATGCAAAAT TTGCTTATTG CAATATTTTC
2901 ATATACAGAC ACTAAAAACA GTTTTCAAAG TCCAGCATTA CGTAACTAAA
2951 GTAAGTAAAA TGATGTGTAT CAACTTGATG GTAAAATATG TAGTTATTTA
3001 AAAAAGCAAT GAACAATTTA GTTTCATGAG AAAATGTTGC CCCCTAAAAG
3051 TAGAACACAT ATGTTACAAC TGCAATAATA CTCTGAATTC ATCTTTCACA
3101 AATAAGAGAC ATGTTAGCAT AGTGATTAAA AGCACAGATA TTGGAGACAA
3151 ACTAACCCAG TTTGAACCCT GGCACTGCCA CGTATAGCAC TGCAGCCTTG
3201 GGAAAGTTAT TTAAACTCAT GGGCTTCAGT TTCAACATCT GTAAAATGGG
3251 CATGTTAACA TTGCCTACCT CATAGGATTA CTGTGAGAAT TTTCTAAGTT
3301 AATATATGTA AAGCAACTTT AAAAAGTGCC TGGCACTTAG TTATTGTTAA
3351 GTAAGTGTCT GCAGATGCAA GTTTGGAAGA GAAAAGCAAA TAAATGAAAA
3401 TCCCTTCCTG TTAAGATGAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
3451 CGGCCGCTCA AGATGAAAAA AAAAAAAAAA AAAAAAAAAA AAAGG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 35 bp to 2137 bp; peptide length: 701
Category: similarity to unknown protein

```
  1 MEEDTDYRIR FSSLCFFNDH VGFHGTIKSS PSDFIVIEID EQGQLVNKTI
 51 DEPIFKISEI QLEPNNFPKK PKLDLQNLSL EDGRNQEVHT LIKYTDGDQN
101 HQSGSEKEDT IVDGTSKCEE KADVLSSFLD EKTHELLNNF ACDVREKWLS
151 KTELIGLPPE FSIGRILDKN QRASLHSAIR QKFPFLVTVG KNSEIVVKPN
201 LEYKELCHLV SEEEAFDFFK YLDAKKENSK FTFKPDTNKD HRKAVHHFVN
251 KKFGNLVETK SFSKMNCSAG NPNVVVTVRF REKAHKRGKR PLSECQEGKV
301 IYTAFTLRKE NLEMFEAIGF LAIKLGVIPS DFSYAGLKDK KAITYQAMVV
351 RKVTPERLKN IEKEIEKKRM NVFNIRSVDD SLRLGQLKGN HFDIVIRNLK
401 KQINDSANLR ERIMEAIENV KKKGFVNYYG PQRFGKGRKV HTDQIGLALL
451 KNEMMKAIKL FLTPEDLDDP VNRAKKYFLQ TEDAKGTLSL MPEFKVRERA
501 LLEALHRFGM TEEGCIQAWF SLPHSMRIFY VHAYTSKIWN EAVSYRLETY
551 GARVVQGDLV CLDEDIDDEN FPNSKIHLVT EEEGSANMYA IHQVVLPVLG
601 YNIQYPKNKV GQWYHDILSR DGLQTCRFKV PTLKLNIPGC YRQILKHPCN
651 LSYQLMEDHD IDVKTKGSHI DETALSLLIS FDLDASCYAT VCLKEIMKHD
701 V
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15g14, frame 2

TREMBL:SPBC1A45P_10 gene: "SPBC1A4.09"; product: "hypothetical protein"; S.pombe chromosome II cosmid c1A4 left hand region 1-26184 bp Originates from chimeric cosmid., N = 3, Score = 511, P = 2.9e-57

PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae), N = 2, Score = 516, P = 7.3e-54

SWISSPROT:YQ4B_CAEEL HYPOTHETICAL 64.6 KD PROTEIN B0024.11 IN CHROMOSOME V., N = 2, Score = 386, P = 2.1e-34

>PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae)
Length = 676

HSPs:

Score = 516 (77.4 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54
Identities = 151/498 (30%), Positives = 245/498 (49%)

```
Query:   191 KNSEIVVKPNLEYKELCHLVSEEEAFDFFK-YLDAKKENSKFTFKPDTNKDHRKAVHHFV 249
             + E V P L  +L + EE+       Y A K +   F+  +K R +H +
```

```
Sbjct:   109 RRQEFNVDPELR-NQLVEIFGEEDVLKIESVYRTANKMETAKNFE---DKSVRTKIHQLL 164

Query:   250 NKKFGNLVETKSFSKMNCSAGNPNVVVTVRFREKAHK-RGKRPLSECQEG-KVIYTAFTL 307
             + F N +E+ +              N         +EK ++ R     +     G   +  FTL
Sbjct:   165 REAFKNELESVTTDTNTFKIARSNRNSRTNKQEKINQTRDANGVENWGYGPSKDFIHFTL 224

Query:   308 RKENLEMFEAIGFLAIKLGVIPSD-FSYAGLKDKKAITYQAMVVRKVTPERLKNIEKEIE 366
             KEN +   EA+  +  KL  +PS    YAG KD++A+T Q + + K+  +RL  + + +
Sbjct:   225 HKENKDTMEAVNVIT-KLLRVPSRVIRYAGTKDRRAVTCQRVSISKIGLDRLNALNRTL- 282

Query:   367 KKRMNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENVKKKGFV 426
             K M + N     D SL LG LKGN F +VIR++     N    +L E +       +++ + GF+
Sbjct:   283 -KGMIIGNYNFSDASLNLGDLKGNEFVVVIRDVTTG-NSEVSLEEIVSNGCKSLSENGFI 340

Query:   427 NYYGPQRFGKGRKVHTDQIGLALLKNEMMKAIKLFLTPEDLDDPVNR-AKKYFLQTEDAK 485
             NY+G QRFG     + T   IG  LL +     KA +L L+ +D    P ++ A+K + +T+DA
Sbjct:   341 NYFGMQRFGTF-SISTHTIGRELLLSNWKKAAELILSDQDNVLPKSKEARKIWAETKDAA 399

Query:   486 GTLSLMPEFKVRERALLEALHRFGMTEEGCIQ--AWFS----LPHSMRIFYVHAYTSKIW 539
             L  MP   + E ALL +L         E+G     A+++    +P ++R  YVHAY  S +W
Sbjct:   400 LALKQMPRQCLAENALLYSLSNQRKEEDGTYSENAYYTAIMKIPRNLRTMYVHAYQSYVW 459

Query:   540 NEAVSYRLETYGARVVQGDLVC-------LDEDIDDENFPNS-------KIHLVTEEEGS 585
             N    S R+E +G ++V GDLV             L    IDDE+F              +    VT+E+
Sbjct:   460 NSIASKRIELHGLKLVVGDLVIDTSEKSPLISGIDDEDFDEDVREAQFIRAKAVTQEDID 519

Query:   586 ANMYAIHQVVLPVLGYNIQYPKNK-VGQWYHDILSRDGLQTCRFKVPTLKLNIPGCYRQI 644
             +   Y +   VVLP   G+++ YP N+ + Q Y DIL   D +     +        ++ G YR +
Sbjct:   520 SVKYTMEDVVLPSPGFDVLYPSNEELKQLYVDILKADNMDPFNMRRKVRDFSLAGSYRTV 579

Query:   645 LKHPCNLSYQLMEDHDIDVKTKGSHID 671
             ++ P +L Y+++     D    +     + +D  .
Sbjct:   580 IQKPKSLEYRIIHYDDPSQQLVNTDLD 606
```

Score = 86 (12.9 bits), Expect = 3.2e-01, Sum P(2) = 2.8e-01
Identities = 40/160 (25%), Positives = 77/160 (48%)

```
Query:    22 GFHGTIKSSPSDFIVIEIDEQGQLVNKTIDEPIFKISEIQLEPNNFPKKPKLDLQNLSLE 81
             GF G IK    +DF+V EID++G++++ T D+   FK+ +    +P       K +++ +   S E
Sbjct:    55 GFRGQIKQRYTDFLVNEIDQEGKVIHLT-DKG-FKMPK---KPQR--SKEEVNAEKES-E 106

Query:    82 DGRNQEVHTLIKYTDGDQNHQSGS--EKEDTI-VDGTSKCEEKADVLSSFLDEKTHELLN 138
             R QE +        D +  +Q         +ED + ++    +     K +     +F D+        ++
Sbjct:   107 AARRQEFNV-----DPELRNQLVEIFGEEDVLKIESVYRTANKMETAKNFEDKSVRTKIH 161

Query:   139 NFACDVREKWLSKTELIGLPPE-FSIGRILDKNQRASLHSAIRQ 181
             +RE + ++ E +      F I R  ++N R +     I Q
Sbjct:   162 QL---LREAFKNELESVTTDTNTFKIARS-NRNSRTNKQEKINQ 201
```

Score = 58 (8.7 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54
Identities = 10/23 (43%), Positives = 17/23 (73%)

```
Query:   676 SLLISFDLDASCYATVCLKEIMK 698
             ++++ F L   S YAT+ L+E+MK
Sbjct:   638 AVVLKFQLGTSAYATMALRELMK 660
```

Pedant information for DKFZphtes3_15g14, frame 2

Report for DKFZphtes3_15g14.2

[LENGTH]	701
[MW]	80700.96
[pI]	7.31
[HOMOL]	PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae) 2e-51
[FUNCAT]	99 unclassified proteins [S. cerevisiae, YOR243c] 8e-53
[BLOCKS]	BL01268C
[BLOCKS]	BL01268B
[BLOCKS]	BL01268A
[SUPFAM]	hypothetical protein HI0701 3e-06
[PROSITE]	MYRISTYL 7
[PROSITE]	AMIDATION 2
[PROSITE]	CAMP_PHOSPHO_SITE 1
[PROSITE]	CK2_PHOSPHO_SITE 16
[PROSITE]	TYR_PHOSPHO_SITE 1
[PROSITE]	PKC_PHOSPHO_SITE 13
[PROSITE]	ASN_GLYCOSYLATION 5
[KW]	Alpha_Beta

```
SEQ    MEEDTDYRIRFSSLCFFNDHVGFHGTIKSSPSDFIVIEIDEQGQLVNKTIDEPIFKISEI
PRD    ccccceeeeeecceeeccccccccceeeeeccccceeeeeeeccccceeeeeecccccceeeeeee

SEQ    QLEPNNFPKKPKLDLQNLSLEDGRNQEVHTLIKYTDGDQNHQSGSEKEDTIVDGTSKCEE
PRD    ccccccccccccccccccccccccccccceeeeccccccccccccccceeeeeeccccccch

SEQ    KADVLSSFLDEKTHELLNNFACDVREKWLSKTELIGLPPEFSIGRILDKNQRASLHSAIR
PRD    hhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhheeeccccccceeeeeeecchhhhhhhhh

SEQ    QKFPFLVTVGKNSEIVVKPNLEYKELCHLVSEEEAFDFFKYLDAKKENSKFTFKPDTNKD
PRD    hhccceeeecccceeeeccccchhhhhhhhhhhhhhhhhhhhhhcccccceeeeeccccccch

SEQ    HRKAVHHFVNKKFGNLVETKSFSKMNCSAGNPNVVVTVRFREKAHKRGKRPLSECQEGKV
PRD    hhhhhhhhhhhhhhhheeeeeeccccceeeeeccccccceeeechhhhhhhhcccccccccccce

SEQ    IYTAFTLRKENLEMFEAIGFLAIKLGVIPSDFSYAGLKDKKAITYQAMVVRKVTPERLKN
PRD    eeeeeeeecccccchhhhhhhhhhhhhhcccccceeeeccccchhhhhhhhheeecccccchhhhh

SEQ    IEKEIEKKRMNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENV
PRD    hhhhhhhhhheeeeeeecccccccccccccccccceeeeeehhhhhccccchhhhhhhhhhhhh

SEQ    KKKGFVNYYGPQRFGKGRKVHTDQIGLALLKNEMMKAIKLFLTPEDLDDPVNRAKKYFLQ
PRD    hhcccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccccchhhhhhhhhh

SEQ    TEDAKGTLSLMPEFKVRERALLEALHRFGMTEEGCIQAWFSLPHSMRIFYVHAYTSKIWN
PRD    hcccchhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhccccchhhhhhhhhhhhhhhh

SEQ    EAVSYRLETYGARVVQGDLVCLDEDIDDENFPNSKIHLVTEEEGSANMYAIHQVVLPVLG
PRD    hhhhhhhhhhcceeeccceeeeccccccccccccccceeeccccccccccccccceeeccccc

SEQ    YNIQYPKNKVGQWYHDILSRDGLQTCRFKVPTLKLNIPGCYRQILKHPCNLSYQLMEDHD
PRD    cccccccccchhhhhhhhhhccccccccccccccccccchhhhhhhhcccchhhhhhhhhcc

SEQ    IDVKTKGSHIDETALSLLISFDLDASCYATVCLKEIMKHDV
PRD    ceeeccccchhhhhhheeeeeeccccccchhhhhhhhhhccc
```

Prosite for DKFZphtes3_15g14.2

PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00001	77->81	ASN_GLYCOSYLATION	PDOC00001
PS00001	266->270	ASN_GLYCOSYLATION	PDOC00001
PS00001	404->408	ASN_GLYCOSYLATION	PDOC00001
PS00001	650->654	ASN_GLYCOSYLATION	PDOC00001
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	237->240	PKC_PHOSPHO_SITE	PDOC00005
PS00005	277->280	PKC_PHOSPHO_SITE	PDOC00005
PS00005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS00005	381->384	PKC_PHOSPHO_SITE	PDOC00005
PS00005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	625->628	PKC_PHOSPHO_SITE	PDOC00005
PS00005	632->635	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	95->99	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	127->131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	150->154	CK2_PHOSPHO_SITE	PDOC00006
PS00006	211->215	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	463->467	CK2_PHOSPHO_SITE	PDOC00006
PS00006	580->584	CK2_PHOSPHO_SITE	PDOC00006
PS00006	668->672	CK2_PHOSPHO_SITE	PDOC00006
PS00007	537->546	TYR_PHOSPHO_SITE	PDOC00007
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	114->120	MYRISTYL	PDOC00008

PS00008	326->332	MYRISTYL	PDOC00008
PS00008	385->391	MYRISTYL	PDOC00008
PS00008	514->520	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00009	287->291	AMIDATION	PDOC00009
PS00009	436->440	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_15g14.2)

DKFZphtes3_15h1

group: testes derived

DKFZphtes3_15h1 encodes a novel 672 amino acid protein with very weak similarity to several proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to Hsp70/Hsp90 organizing protein

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2277 bp
Poly A stretch at pos. 2252, polyadenylation signal at pos. 2226

```
   1 AAACCAGATA GAGGTTCTCC AGCTTTTCTT TGATTGTCTC TGCTTTAGCG
  51 TCTCTAAATC CGGTCACCAT GTCGGACCCC GAAGGCGAGA CCTTGCGAAG
 101 CACCTTTCCC TCTTATATGG CCGAAGGCGA GCGGCTCTAC CTGTGCGGGG
 151 AATTTTCTAA AGCCGCGCAG AGCTTCAGCA ACGCTCTTTA CCTTCAGGAT
 201 GGAGACAAGA ACTGCCTGGT TGCTCGCTCA AAGTGCTTCC TGAAGATGGG
 251 AGACTTGGAG AGATCCCTGA AGGATGCTGA GGCTTCGCTC CAGAGTGACC
 301 CAGCTTTCTG TAAGGGGATT TTGCAAAAGG CTGAGACACT GTACACCATG
 351 GGAGACTTTG AGTTTGCCTT GGTATTCTAT CATCGAGGCT ACAAGCTGAG
 401 GCCTGATCGG GAATTCAGAG TTGGCATTCA GAAAGCCCAG GAAGCCATCA
 451 ACAACTCAGT GGGAAGTCCT TCTTCCATTA AGCTGGAGAA CAAAGGGGAC
 501 CTCTCCTTCT TAAGCAAGCA GGCTGAGAAT ATAAAAGCCC AGCAGAAGCC
 551 TCAGCCCATG AAACACCTCT TACACCCCAC CAAGGGAGAG CCCAAGTGGA
 601 AGGCCTCGCT CAAGAGTGAG AAGACTGTCC GCCAGCTTCT GGGGGAGCTC
 651 TACGTGGACA AAGAGTATTT GGAGAAGCTC CTATTGGATG AAGACCTGAT
 701 CAAAGGCACC ATGAAGGGCG GCCTGACTGT GGAGGACCTC ATCATGACGG
 751 GCATCAACTA CCTGGATACT CACAGCAACT TCTGGAGGCA GCAGAAGCCG
 801 ATCTACGCCA GGGAGCGGGA CCGGAAGCTG ATGCAAGAGA AATGGCTGCG
 851 GGACCACAAA CGCCGTCCCT CACAGACAGC CCATTACATC CTCAAGAGCC
 901 TGGAGGACAT TGATATGTTG CTCACAAGTG GCAGTGCTGA AGGGAGTCTT
 951 CAGAAAGCTG AGAAAGTGCT GAAGAAGGTA CTGGAATGGA ACAAGGAAGA
1001 GGTACCCAAC AAGGATGAAC TGGTTGGAAA CTTGTATAGC TGCATAGGGA
1051 ATGCCCAGAT TGAGCTGGGG CAGATGGAGG CAGCCCTGCA GAGCCACAGA
1101 AAGGACCTGG AGATCGCCAA GGAATATGAC CTTCCTGATG CAAAATCGAG
1151 AGCCCTTGAC AACATTGGCA GAGTTTTTGC CAGAGTTGGG AAATTCCAGC
1201 AAGCCATTGA CACGTGGGAA GAAAAGATCC CTCTGGCAAA AACCACCCTG
1251 GAGAAGACCT GGCTGTTCCA CGAGATCGGC CGCTGCTACT TGGAGCTGGA
1301 CCAGGCCTGG CAGGCCCAGA ATTATGGCGA GAAGTCCCAG CAGTGTGCCG
1351 AGGAGGAAGG GGACATTGAG TGGCAACTGA ATGCCAGTGT TCTGGTGGCC
1401 CAGGCACAAG TGAAGCTGAG AGACTTCGAG TCAGCCGTGA ACAATTTTGA
1451 GAAGGCCCTG GAGAGAGCAA AGCTTGTGCA TAACAACGAG GCGCAGCAGG
1501 CCATCATCAG TGCCTTGGAC GATGCCAACA AGGGTATCAT CAGAGAACTG
1551 AGGAAAAACCA ACTACGTGGA GAATCTCAAA GAAAAAAGCG AGGGAGAAGC
1601 TTCACTGTAT GAAGATAGAA TAATAACAAG AGAGAAGGAC ATGAGGAGAG
1651 TGAGAGATGA GCCCGAGAAG GTGGTGAAGC AGTGGGACCA TAGTGAGGAT
1701 GAGAAAGAGA CAGATGAGGA CGATGAGGCT TTTGGGGAAG CTCTGCAGAG
1751 CCCAGCAAGC GGAAAGCAGA GTGTGGAAGC AGGAAAAGCC AGAAGCGATT
1801 TGGGAGCAGT TGCCAAGGGC CTGTCAGGAG AATTAGGCAC AAGATCAGGA
1851 GAAACAGGCA GGAAGCTACT AGAAGCTGGC AGAAGAGAGT CAAGAGAAAT
1901 TTATAGGAGG CCTTCGGGAG AATTAGAGCA AAGACTCTCA GGAGAATTCA
1951 GCAGACAGGA ACCAGAAGAA CTAAAGAAAC TTTCAGAAGT GGGCAGAAGA
2001 GAGCCAGAAG AACTGGGAAA AACACAATTT GGAGAAATAG GAGAAACGAA
2051 AAAAACAGGA AATGAGATGG AAAAGGAATA TGAATGAAGC CATCGGTAGA
2101 GATGAGGATC AGGAAGCTGG TGTTCAGAGG GATCATGGGA TTTTATTAAA
2151 CTGGATTTTC AAGCGATTTG TCTGTTATAG GAAAAATGAG GGTTTTACTT
2201 CTGCTGCTTT CCATCACTAT TTTGCCATTA AATAGGTGTC TTTCACTCTT
2251 GCAAAAAAAA AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 69 bp to 2084 bp; peptide length: 672
Category: similarity to known protein

```
  1 MSDPEGETLR STFPSYMAEG ERLYLCGEFS KAAQSFSNAL YLQDGDKNCL
 51 VARSKCFLKM GDLERSLKDA EASLQSDPAF CKGILQKAET LYTMGDFEFA
101 LVFYHRGYKL RPDREFRVGI QKAQEAINNS VGSPSSIKLE NKGDLSFLSK
151 QAENIKAQQK PQPMKHLLHP TKGEPKWKAS LKSEKTVRQL LGELYVDKEY
201 LEKLLLDEDL IKGTMKGGLT VEDLIMTGIN YLDTHSNFWR QQKPIYARER
251 DRKLMQEKWL RDHKRRPSQT AHYILKSLED IDMLLTSGSA EGSLQKAEKV
301 LKKVLEWNKE EVPNKDELVG NLYSCIGNAQ IELGQMEAAL QSHRKDLEIA
351 KEYDLPDAKS RALDNIGRVF ARVGKFQQAI DTWEEKIPLA KTTLEKTWLF
401 HEIGRCYLEL DQAWQAQNYG EKSQQCAEEE GDIEWQLNAS VLVAQAQVKL
451 RDFESAVNNF EKALERAKLV HNNEAQQAII SALDDANKGI IRELRKTNYV
501 ENLKEKSEGE ASLYEDRIIT REKDMRRVRD EPEKVVKQWD HSEDEKETDE
551 DDEAFGEALQ SPASGKQSVE AGKARSDLGA VAKGLSGELG TRSGETGRKL
601 LEAGRRESRE IYRRPSGELE QRLSGEFSRQ EPEELKKLSE VGRREPEELG
651 KTQFGEIGET KKTGNEMEKE YE
```

BLASTP hits

Entry AF039202_1 from database TREMBL:
product: "Hsp70/Hsp90 organizing protein"; Cricetulus griseus Hsp70/Hsp90 organizing protein mRNA, complete cds.
Score = 149, P = 5.3e-07, identities = 42/160, positives = 74/160

Entry AI09782_1 from database TREMBL:
product: "myosin heavy chain"; Argopecten irradians myosin heavy chain mRNA, complete cds.
Score = 155, P = 6.1e-07, identities = 140/623, positives = 256/623

Entry S56658 from database PIR:
stress-induced protein stil - soybean
Score = 156, P = 9.7e-08, identities = 41/153, positives = 72/153

Alert BLASTP hits for DKFZphtes3_15h1, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15h1, frame 3

Report for DKFZphtes3_15h1.3

```
[LENGTH]     672
[MW]         76655.61
[pI]         5.49
[HOMOL]      PIR:S56658 stress-induced protein stil - soybean 6e-10
[SUPFAM]     tetratricopeptide repeat homology 1e-07
[PROSITE]    MYRISTYL        7
[PROSITE]    AMIDATION       3
[PROSITE]    CAMP_PHOSPHO_SITE      4
[PROSITE]    CK2_PHOSPHO_SITE       15
[PROSITE]    TYR_PHOSPHO_SITE       1
[PROSITE]    PKC_PHOSPHO_SITE       11
[PROSITE]    ASN_GLYCOSYLATION      2
[KW]         All_Alpha
[KW]         LOW_COMPLEXITY     4.76 %

SEQ  MSDPEGETLRSTFPSYMAEGERLYLCGEFSKAAQSFSNALYLQDGDKNCLVARSKCFLKM
SEG  ............................................................
PRD  cccccccceeeccccccccccccccccchhhhhhhhhhhhhcccccceeehhhhhhhhhh

SEQ  GDLERSLKDAEASLQSDPAFCKGILQKAETLYTMGDFEFALVFYHRGYKLRPDREFRVGI
SEG  ............................................................
PRD  hcchhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhh
```

```
SEQ     QKAQEAINNSVGSPSSIKLENKGDLSFLSKQAENIKAQQKPQPMKHLLHPTKGEPKWKAS
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhcchhhhhhhhchhhhhhhcccchhhhhhcccccccchhhh

SEQ     LKSEKTVRQLLGELYVDKEYLEKLLLDEDLIKGTMKGGLTVEDLIMTGINYLDTHSNFWR
SEG     ............xxxxxxxxxxxxxxxxxx..............................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhccccccccccccccc

SEQ     QQKPIYARERDRKLMQEKWLRDHKRRPSQTAHYILKSLEDIDMLLTSGSAEGSLQKAEKV
SEG     ............................................................
PRD     cchhhhhhhhhhhhhhhhhhccccccchhhhhhhhhhhheeeeecccccchhhhhhhhhh

SEQ     LKKVLEWNKEEVPNKDELVGNLYSCIGNAQIELGQMEAALQSHRKDLEIAKEYDLPDAKS
SEG     ............................................................
PRD     hhhhhhhcccccccccccceeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhcccccchh

SEQ     RALDNIGRVFARVGKFQQAIDTWEEKIPLAKTTLEKTWLFHEIGRCYLELDQAWQAQNYG
SEG     ............................................................
PRD     hhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhh

SEQ     EKSQQCAEEEGDIEWQLNASVLVAQAQVKLRDFESAVNNFEKALERAKLVHNNEAQQAII
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhh

SEQ     SALDDANKGIIRELRKTNYVENLKEKSEGEASLYEDRIITREKDMRRVRDEPEKVVKQWD
SEG     ...........................................................x
PRD     hhhhccchhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhccceeeeecc

SEQ     HSEDEKETDEDDEAFGEALQSPASGKQSVEAGKARSDLGAVAKGLSGELGTRSGETGRKL
SEG     xxxxxxxxxxxxx...............................................
PRD     ccccccccccchhhhhhcccccccccchhhhccccccceeeeeccccccccccccccchhh

SEQ     LEAGRRESREIYRRPSGELEQRLSGEFSRQEPEELKKLSEVGRREPEELGKTQFGEIGET
SEG     ............................................................
PRD     hhhccccceeeeccccccchhhhccccccchhhhhhhhhhhcccccccccccccccccccc

SEQ     KKTGNEMEKEYE
SEG     ............
PRD     cccccccccccc
```

Prosite for DKFZphtes3_15h1.3

PS00001	128->132	ASN_GLYCOSYLATION	PDOC00001
PS00001	438->442	ASN_GLYCOSYLATION	PDOC00001
PS00004	265->269	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	605->609	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	613->617	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	636->640	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	8->11	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00005	136->139	PKC_PHOSPHO_SITE	PDOC00005
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00005	183->186	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	214->217	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	564->567	PKC_PHOSPHO_SITE	PDOC00005
PS00005	596->599	PKC_PHOSPHO_SITE	PDOC00005
PS00005	660->663	PKC_PHOSPHO_SITE	PDOC00005
PS00006	2->6	CK2_PHOSPHO_SITE	PDOC00006
PS00006	66->70	CK2_PHOSPHO_SITE	PDOC00006
PS00006	93->97	CK2_PHOSPHO_SITE	PDOC00006
PS00006	171->175	CK2_PHOSPHO_SITE	PDOC00006
PS00006	220->224	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00006	382->386	CK2_PHOSPHO_SITE	PDOC00006
PS00006	392->396	CK2_PHOSPHO_SITE	PDOC00006
PS00006	481->485	CK2_PHOSPHO_SITE	PDOC00006
PS00006	507->511	CK2_PHOSPHO_SITE	PDOC00006
PS00006	512->516	CK2_PHOSPHO_SITE	PDOC00006
PS00006	542->546	CK2_PHOSPHO_SITE	PDOC00006
PS00006	548->552	CK2_PHOSPHO_SITE	PDOC00006
PS00006	628->632	CK2_PHOSPHO_SITE	PDOC00006
PS00006	663->667	CK2_PHOSPHO_SITE	PDOC00006
PS00007	506->515	TYR_PHOSPHO_SITE	PDOC00007
PS00008	119->125	MYRISTYL	PDOC00008
PS00008	132->138	MYRISTYL	PDOC00008
PS00008	213->219	MYRISTYL	PDOC00008

PS00008	288->294	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	334->340	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00009	596->600	AMIDATION	PDOC00009
PS00009	603->607	AMIDATION	PDOC00009
PS00009	641->645	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_15h1.3)

DKFZphtes3_15i5
----

group: cell structure and motility

DKFZphtes3_15i5 encodes a novel 717 amino acid protein with similarity to radial spokehead proteins.

The novel protein is similar to the Chlamydomonas reinhardtii radial spokehead protein of flagella or axoneme and to the Strongylocentrotus purpuratus sea urchin spermatozoa protein p63. This protein is important for the maintenance of a planar form of sperm flagellar beating. In addition, the novel protein contains a transferrin signature 1 for iron-binding. The new protein seems to be a part of the human radial spoke heads in spermatozoa.

BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in modulating the structure of the human spermatozoa radia spoke head and modulation of sperm motility in men.

strong similarity to "radial spokehead" proteins

complete cDNA, complete cds, 1 EST hit (from a testis library)
"radial spokehead" part of flagella in Chlamydomona, this protein seems to be part of the sperm motor or tail

Sequenced by GBF

Locus: unknown

Insert length: 2478 bp
Poly A stretch at pos. 2452, polyadenylation signal at pos. 2433

```
   1 CACCCTGGCC CGCTCCCCGC GCCCTCCACG GGTAACGGCC CCCTCTCTCG
  51 GTGCTCAGAA ACCGGCGGTG TCGACAGGTG GCTCTCGCTT GGCCTCCTTG
 101 TCTGCAAGCC TTTCTCCTAG AGATCTGTGC CTCCTGGCGA ACCATGGGAG
 151 ACCTGCCGCC CTACCCTGAG CGCCCTGCCC AGCAGCCTCC GGGCCGGAGG
 201 ACTTCTCAGG CCTCCCAGAG GCGGCACAGT CGGGACCAAG CTCAGGCCCT
 251 GGCAGCGGAC CCCGAGGAGA GGCAGCAGAT ACCTCCAGAC GCCCAGCGAA
 301 ACGCCCCTGG TTGGTCACAG AGGGGCAGCC TGTCCCAACA GGAGAACTTG
 351 CTGATGCCCC AGGTCTTCCA GGCTGAGGAA GCCCGGCTGG GTGGCATGGA
 401 GTACCCATCT GTGAACACGG GCTTTCCCTC AGAGTTCCAG CCTCAGCCTT
 451 ACTCTGATGA AAGCAGGATG CAGGTCGCCG AGCTCACCAC CAGCCTAATG
 501 CTGCAGCGGC TCCAGCAGGG CCAAAGCAGC CTGTTCCAGC AACTGGACCC
 551 CACCTTCCAG GAGCCCCCAG TCAACCCCTT GGGCCAGTTC AACCTCTACC
 601 AGACAGACCA GTTCTCTGAA GGTGCCCAGC ACGGGCCTTA CATAAGGGAT
 651 GACCCTGCCC TTCAGTTCTT GCCCTCTGAG CTGGGCTTCC CACACTACAG
 701 TGCCCAGGTG CCTGAGCCCG AGCCTCTGGA GCTGGCCGTG CAGAACGCCA
 751 AGGCCTACCT GCTGCAGACC AGCATCAATT GCGACCTCAG CCTGTACGAG
 801 CACCTGGTAA ATCTGCTGAC CAAGATCCTG AACCAGCGGC CTGAGGACCC
 851 CTTGTCTGTC CTGGAGTCTC TGAACCGCAC CACGCAGTGG GAGTGGTTCC
 901 ACCCCAAGCT GGACACGCTG CGGGACGACC CCGAGATGCA GCCCACCTAC
 951 AAGATGGCGG AGAAACAGAA GGCGCTGTTC ACCCGGAGTG GAGGCGGCAC
1001 TGAAGGCGAA CAGGAGATGG AGGAGGAGGT GGGGGAGACA CCAGTGCCCA
1051 ACATCATGGA GACTGCCTTC TACTTCGAGC AGGCCGGCGT CGGCCTGAGC
1101 TCGGACGAGA GCTTCCGCAT TTTCCTGGCC ATGAAACAGC TGGTGGAGCA
1151 GCAGCCCATC CACACCTGTC GCTTCTGGGG CAAGATCCTG GGAATCAAAC
1201 GCAGCTACCT GGTGGCCGAG GTGGAATTCC GGGAGGGCGA GGAGGAGGCA
1251 GAGGAGGAGG AGGTGGAGGA GATGACGGAA GGTGGCGAGG TCATGGAGGC
1301 GCACGGCGAG GAGGAGGGCG AGGAGGACGA GGAGAAGGCC GTGGACATCG
1351 TCCCTAAGTC CGTATGGAAG CCGCCGCCCG TGATCCCCAA GGAGGAGAGC
1401 CGCTCAGGCG CCAACAAGTA CCTGTACTTT GTGTGCAACG AGCCGGGCCT
1451 GCCATGGACG CGGCTGCCCC ACGTCACTCC AGCCCAGATC GTGAACGCCC
1501 GAAAGATCAA GAAGTTCTTC ACAGGCTACC TGGACACGCC AGTCGTCAGC
1551 TACCCACCCT TCCCGGGCAA CGAGGCCAAC TACCTGCGGG CCCAGATAGC
1601 CCGCATCTCG GCCGCCACGC AGGTCAGCCC GCTGGGCTTC TACCAGTTTA
1651 GTGAGGAGGA GGGCGACGAG GAGGAGGAAG GTGGTGCTGG GCGCGACTCC
1701 TACGAGGAGA ACCCGGACTT CGAGGGCATC CCCGTGCTGG AGCTGGTCGA
1751 CTCCATGGCC AACTGGGTGC ATCACACACA GCACATCCTG CCGCAGGGCC
1801 GCTGCACTTG GGTGAACCCT TTGCAGAAGA CAGAGGAGGA GGAGGACCTG
1851 GGGGAGGAGG AAGAGAAGGC AGATGAGGGG CCAGAGGAGG TGGAGCAGGA
1901 GGTTGGCCCC CCACTGCTAA CGCCACTTTC AGAAGATGCA GAAATCATGC
1951 ACCTGGCACC CTGGACCACC CGCCTGTCCT GCAGCCTCTG CCCGCAGTAC
2001 TCAGTGGCCG TTGTGCGCTC CAACCTCTGG CCCGGGGCCT ATGCCTATGC
2051 CAGTGGCAAA AAGTTTGAGA ACATCTACAT CGGCTGGGGT CACAAGTACA
2101 GCCCCGAGAG CTTCAACCCG GCCCTGCCAG CCCCCATTCA ACAAGAGTAC
2151 CCCAGTGGCC CAGAGATCAT GGAGATGAGT GACCCCACAG TGGAAGAGGA
2201 GCAGGCTCTG AAAGCAGCCC AGGAACAAGC CCTGGGAGCC ACAGAGGAGG
2251 AGGAGGAGGG CGAGGAGGAG GAGGAGGGCG AGGAGACAGA TGACTGAGGC
```

```
2301 CCACCCTCTA GCCACTTTCC CCAAGCAGGT AGATAGCAAA TTTCCCCTTA
2351 GAGGTAGTTA GCATGGATTA TATTTTCACT ATGTGCTTCC TGTCCCCAGA
2401 GGGCAGGGAT AGAAAAGGAA GGCAACTGCT TCAAATAAAA TTCCTCCACG
2451 GCATTAAAAA AAAAAAAAAA AAAAAAAG
```

BLAST Results

No BLAST result

Medline entries

86251010:
Molecular cloning and expression of flagellar radial spoke and dynein genes of Chlamydomona

81142496:
Radial spokes of Chlamydomonas flagella: polypeptide composition and phosphorylation of stalk components.

9450971:
Molecular cloning and characterization of a radial spoke head protein of sea urchin sperm axonemes: involvement of the protein in the regulation of sperm motility.

Peptide information for frame 3

ORF from 144 bp to 2294 bp; peptide length: 717
Category: strong similarity to known protein

```
  1 MGDLPPYPER PAQQPPGRRT SQASQRRHSR DQAQALAADP EERQQIPPDA
 51 QRNAPGWSQR GSLSQQENLL MPQVFQAEEA RLGGMEYPSV NTGFPSEFQP
101 QPYSDESRMQ VAELTTSLML QRLQQGQSSL FQQLDPTFQE PPVNPLGQFN
151 LYQTDQFSEG AQHGPYIRDD PALQFLPSEL GFPHYSAQVP EPEPLELAVQ
201 NAKAYLLQTS INCDLSLYEH LVNLLTKILN QRPEDPLSVL ESLNRTTQWE
251 WFHPKLDTLR DDPEMQPTYK MAEKQKALFT RSGGGTEGEQ EMEEEVGETP
301 VPNIMETAFY FEQAGVGLSS DESFRIFLAM KQLVEQQPIH TCRFWGKILG
351 IKRSYLVAEV EFREGEEEAE EEEVEEMTEG GEVMEAHGEE EGEEDEEKAV
401 DIVPKSVWKP PPVIPKEESR SGANKYLYFV CNEPGLPWTR LPHVTPAQIV
451 NARKIKKFFT GYLDTPVVSY PPFPGNEANY LRAQIARISA ATQVSPLGFY
501 QFSEEEGDEE EEGGAGRDSY EENPDFEGIP VLELVDSMAN WVHHTQHILP
551 QGRCTWVNPL QKTEEEEDLG EEEEKADEGP EEVEQEVGPP LLTPLSEDAE
601 IMHLAPWTTR LSCSLCPQYS VAVVRSNLWP GAYAYASGKK FENIYIGWGH
651 KYSPESFNPA LPAPIQQEYP SGPEIMEMSD PTVEEEQALK AAQEQALGAT
701 EEEEEGEEEE EGEETDD
```

BLASTP hits

Entry U73123_1 from database TREMBL:
product: "radial spokehead"; Strongylocentrotus purpuratus radial spokehead mRNA, complete cds.
Score = 1604, P = 7.4e-165, identities = 303/523, positives = 395/523

Entry B44498 from database PIR:
radial spoke protein 6 - Chlamydomonas reinhardtii
Score = 386, P = 3.4e-45, identities = 105/264, positives = 138/264

Alert BLASTP hits for DKFZphtes3_15i5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15i5, frame 3

Report for DKFZphtes3_15i5.3

[LENGTH]	717
[MW]	80913.61
[pI]	4.36

[HOMOL] TREMBL:U73123_1 product: "radial spokehead"; Strongylocentrotus purpuratus radial spokehead mRNA, complete cds. 1e-130
[PROSITE] TRANSFERRIN_1 1
[PROSITE] MYRISTYL 5
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 14
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] GLYCOSAMINOGLYCAN 1
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 21.48 %

```
SEQ  MGDLPPYPERPAQQPPGRRTSQASQRRHSRDQAQALAADPEERQQIPPDAQRNAPGWSQR
SEG  ....xxxxxxxxxxxx............................................
PRD  cccccccccccccccccccchhhhhhhhhhhhhhhhhcccccccccccccccccccccc

SEQ  GSLSQQENLLMPQVFQAEEARLGGMEYPSVNTGFPSEFQPQPYSDESRMQVAELTTSLML
SEG  ........................................................xxxx
PRD  cccchhhhhhhhhhhhhhhhhhcccccccccccccccccccccchhhhhhhhhhhhhhh

SEQ  QRLQQGQSSLFQQLDPTFQEPPVNPLGQFNLYQTDQFSEGAQHGPYIRDDPALQFLPSEL
SEG  xxxxxxxxxxxxx...............................................
PRD  hhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  GFPHYSAQVPEPEPLELAVQNAKAYLLQTSINCDLSLYEHLVNLLTKILNQRPEDPLSVL
SEG  ............................................................
PRD  cccccccccccccchhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhcccccchhhh

SEQ  ESLNRTTQWEWFHPKLDTLRDDPEMQPTYKMAEKQKALFTRSGGGTEGEQEMEEEVGETP
SEG  .............................................xxxxxxxxxxxxxxx..
PRD  hhhchhhhhcccccccccccccccchhhhhhhhhhhhhhhcccccchhhhhhhhhcccc

SEQ  VPNIMETAFYFEQAGVGLSSDESFRIFLAMKQLVEQQPIHTCRFWGKILGIKRSYLVAEV
SEG  .........................................................xxx
PRD  ccchhhhhhhhhhccccccchhhhhhhhhhhhhhhhhcchhhhhhhhhccchhhhhhhh

SEQ  EFREGEEEAEEEEVEEMTEGGEVMEAHGEEEGEEDEEKAVDIVPKSVWKPPPVIPKEESR
SEG  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx......................
PRD  hhhhhhhhhhhhhhhhhhcccccccccccccchhhhheeeeccccccccccccccccccc

SEQ  SGANKYLYFVCNEPGLPWTRLPHVTPAQIVNARKIKKFFTGYLDTPVVSYPPFPGNEANY
SEG  ............................................................
PRD  ccccceeeeeeccccccccccccccchhhhhhhhhhhhhcccccccccccccccchhhh

SEQ  LRAQIARISAATQVSPLGFYQFSEEEGDEEEEGGAGRDSYEENPDFEGIPVLELVDSMAN
SEG  ......................xxxxxxxxxxxxx.........................
PRD  hhhhhhhhhhhhccccccceeeecccccccccccccccccccccccceeeecchhhh

SEQ  WVHHTQHILPQGRCTWVNPLQKTEEEEDLGEEEEKADEGPEEVEQEVGPPLLTPLSEDAE
SEG  ......................xxxxxxxxxxxxxxxxxxxxxxxxxxx..........
PRD  hhhccccccccceeechhhhhhhhhccccchhhhhccccccccccccccccccccccc

SEQ  IMHLAPWTTRLSCSLCPQYSVAVVRSNLWPGAYAYASGKKFENIYIGWGHKYSPESFNPA
SEG  ............................................................
PRD  cccccccccccccccccccceeeeeeccccceeeecccccceeeeeecccccccccccc

SEQ  LPAPIQQEYPSGPEIMEMSDPTVEEEQALKAAQEQALGATEEEEEGEEEEEGEETDD
SEG  ......................xxxxxxxxxxxxxx...xxxxxxxxxxxxxx...
PRD  ccccccccccccceeeecccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccc
```

Prosite for DKFZphtes3_15i5.3

PS00001	244->248	ASN_GLYCOSYLATION	PDOC00001
PS00002	282->286	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	18->22	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	26->30	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	637->640	PKC_PHOSPHO_SITE	PDOC00005
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006

PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	286->290	CK2_PHOSPHO_SITE	PDOC00006
PS00006	319->323	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	671->675	CK2_PHOSPHO_SITE	PDOC00006
PS00006	682->686	CK2_PHOSPHO_SITE	PDOC00006
PS00006	700->704	CK2_PHOSPHO_SITE	PDOC00006
PS00007	639->646	TYR_PHOSPHO_SITE	PDOC00007
PS00008	284->290	MYRISTYL	PDOC00008
PS00008	315->321	MYRISTYL	PDOC00008
PS00008	350->356	MYRISTYL	PDOC00008
PS00008	435->441	MYRISTYL	PDOC00008
PS00008	475->481	MYRISTYL	PDOC00008
PS00009	16->20	AMIDATION	PDOC00009
PS00009	637->641	AMIDATION	PDOC00009
PS00205	619->628	TRANSFERRIN_1	PDOC00182

(No Pfam data available for DKFZphtes3_15i5.3)

DKFZphtes3_15j18

group: testes derived

DKFZphtes3_15j18 encodes a novel 148 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 905 bp
Poly A stretch at pos. 839, polyadenylation signal at pos. 815

```
  1 GTGATTCATA TGCTTCCATA GCAGGTGTCT GCTTCTGAGC CAAGCTCCCA
 51 GGGCAGCGGA GCAGGCACCA ACCAGCATCC CAGGGGAGGG CACAGCTTGT
101 CCAGCTGGGA TGTTTGGGTG CCCTGTGAGA TGCCCCAAGC CACCAACCCA
151 GCTTATCTCA GGAGAAGCCT CGGCGGCCCG TCTGCCGGCC TGGAGAGATG
201 TGCTACAGCA GCCGGGGGTG GGGGGAGAGG GTGGGCTTAG AATCTCTTGG
251 CAGGGAGCCC CCAAGAGCAG GGTGAGACCT GCCTTCATTT CACCTGTCCC
301 CTTCACAGTT CTGCAAAGCC AGCATTATCA TCCCTTTTCA GAAGGAGTGG
351 GCACTCAGGT GGAATGCCTC ACCCCAGTCC TGCGGCTGGA AAGCGATATG
401 GCCAGGACTG CACCCCACCC CTCATCCCTG CACCCCTTCC CTGCCTGGGA
451 TTCCTCCAGC CCTGTGCACT GTGGAGCGCC TCTGCCTTCC GCTCATGGAG
501 GTTTCCCAAG GGCACGCGCT GAGGGCAGCT GGTCTCAGCC TGGGGCCGGG
551 TCCTAGTAAC TGTCTCTCTT TGCTTTCCAG CCAGTGTTTT GGGGTTTGAA
601 GTTGGAATCT TCAGCTACTG TCAAGAACAG CCACAAAAAT GTGTCACGAT
651 CAAGATCTTT GAGAGTCCAC CAATCAGGAG GCGTCTGTGA CAGTCGCTGT
701 CTTCTCAGAA CAGAATCCAC ACCCAGGATT CAACCCAAAT GATTTCTCAT
751 CAGGTGATTC TTGGTTGTAG CAAAGTTCAT GTGAATGTGG GTGAGTTTCT
801 GTTATGAATG TGGTCAATAA ATGTTATTTG TGAAACTCTA AAAAAAAAAA
851 AAAAAAAAAG GGCGGCCGCT CTAGAGGATC CAAGCTTACG TACGCGAAAA
901 AAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 110 bp to 553 bp; peptide length: 148
Category: putative protein

```
  1 MFGCPVRCPK PPTQLISGEA SAARLPAWRD VLQQPGVGGE GGLRISWQGA
 51 PKSRVRPAFI SPVPFTVLQS QHYHPFSEGV GTQVECLTPV LRLESDMART
101 APHPSSLHPF PAWDSSSPVH CGAPLPSAHG GFPRARAEGS WSQPGAGS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15j18, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15j18, frame 2

Report for DKFZphtes3_15j18.2

[LENGTH]	148	
[MW]	15665.78	
[pI]	8.91	
[PROSITE]	MYRISTYL	3
[PROSITE]	CK2_PHOSPHO_SITE	1
[KW]	Irregular	

```
SEQ    MFGCPVRCPKPPTQLISGEASAARLPAWRDVLQQPGVGGEGGLRISWQGAPKSRVRPAFI
PRD    cccccccccccccccccccccccchhhhhhhcccccccccccceeeeecccccccccccccc

SEQ    SPVPFTVLQSQHYHPFSEGVGTQVECLTPVLRLESDMARTAPHPSSLHPFPAWDSSSPVH
PRD    cccceeeeeccccccccccccccccccchhhhhhhcccccccccccccccccccccccc

SEQ    CGAPLPSAHGGFPRARAEGSWSQPGAGS
PRD    cccccccccccccccccccccccccccc
```

Prosite for DKFZphtes3_15j18.2

PS00006	82->86	CK2_PHOSPHO_SITE	PDOC00006
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	42->48	MYRISTYL	PDOC00008
PS00008	49->55	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_15j18.2)

DKFZphtes3_15j3
----

group: nucleic acid management

DKFZphtes3_15j3 encodes a novel 743 amino acid protein with similarity to proteins with unknown function.

The novel protein contains a RNA recognition motif, predicted by Pfam and therefore binds to RNA. The protein is similar to YGR276c, a ribonuclease H of S. cerevisiae. Thus, the protein seems to a new RNA-modificating protein.

The new protein can find application in modulating the RNA metabolism in human cells and as a tool for biotechnologic manipulations.

"44M2.3"; product, differences to genmodel, similarity to ribonuclease H

complete cDNA, complete cds, EST hits
YGR276c = ribonuclease H
differences to genmodel of 44M2.3

Sequenced by GBF

Locus: /map="16p11.2"

Insert length: 2695 bp
Poly A stretch at pos. 2601, polyadenylation signal at pos. 2579

```
   1 GCGGTTGTTG TTGGCAGCTG TGGCTAAGGA GGGGAGAACC TCTGCTCCCC
  51 GCCCGTCTTC TCTTCTGCGT TTCCCGGGCT AGGGGGCGTG GGGAGTGGTT
 101 TTAGGCGGCG AAGCCGCTCG GCAGCACCTT CCTTCTTTGC CAGGCAGACG
 151 CCCGTTGTAG CCGTTGGGGA ACCGTTGAGA ATCCGCCATG GAGCCAGAGA
 201 GGGAAGGGAC CGAGAGACAC CCCAGGAAGG TCAGGGAAAG CAGGCAGGCC
 251 CCAAATAAGC TGGTCGGGGC AGCTGAGGCG ATGAAAGCCG GTTGGGATCT
 301 CGAGGAGAGT CAGCCCGAGG CCAAGAAAGC CCGCTTATCT ACCATTTTAT
 351 TTACTGACAA CTGTGAAGTA ACCCATGACC AGCTGTGTGA ATTGCTGAAG
 401 TATGCAGTTC TGGGCAAATC CAATGTTCCA AAACCCAGCT GGTGCCAGCT
 451 TTTTCATCAA AACCACCTAA ACAACGTAGT GGTTTTTGTT CTGCAGGGAA
 501 TGAGTCAGCT ACACTTTTAC AGGTTCTATT TGGAGTTTGG ATGTCTTCGA
 551 AAAGCATTCA GACATAAATT CCGCTTGCCT CCACCATCAT CTGATTTTCT
 601 AGCTGATGTT GTTGGGCTAC AAACTGAACA AAGAGCTGGA GATCTGCCCA
 651 AGACAATGGA AGGGCCTTTA CCTTCTAATG CAAAAGCCGC CATCAACCTT
 701 CAGGATGATC CCATCATTCA AAAGTATGGC TCTAAGAAAG TGGGCTTGAC
 751 CAGATGCCTT CTGACAAAGG AGGAAATGAG AACGTTTCAC TTTCCATTAC
 801 AAGGTTTTCC TGATTGTGAA AACTTTTTAC TTACCAAATG TAATGGTTCT
 851 ATAGCAGACA ATAGTCCTCT CTTTGGACTT GACTGTGAAA TGTGCCTCAC
 901 ATCCAAGGGG AGAGAGCTAA CACGCATCTC ACTGGTTGCT GAAGGAGGCT
 951 GCTGTGTTAT GGATGAACTG GTCAAACCTG AAAACAAGAT TCTGGACTAC
1001 CTCACCAGCT TTTCGGGAAT CACGAAGAAG ATTCTTAACC CAGTGACGAC
1051 CAAACTCAAA GATGTACAGA GGCAGTTAAA AGCACTGCTT CCTCCTGATG
1101 CTGTGTTAGT GGGCCACTCC TTAGATTTGG ATCTCAGAGC ACTGAAAATG
1151 ATACATCCAT ATGTTATTGA TACATCGTTG CTTTATGTCA GAGAGCAGGG
1201 CAGAAGATTT AAGCTCAAGT TCTTAGCCAA AGTTATTTTG GGGAAGGATA
1251 TACAGTGTCC AGACAGACTT GGTCATGATG CCACAGAAGA TGCTAGAACA
1301 ATCCTTGAAT TGGCTCGGTA TTTCCTTAAG CATGGCCCAA AAAAGATTGC
1351 AGAACTAAAT CTAGAAGCAC TAGCTAATCA CCAAGAAATA CAAGCAGCAG
1401 GCCAAGAGCC TAAAAACACA GCAGAAGTAC TTCAGCACCC AAACACAAGT
1451 GTTTTAGAAT GCTTGGATTC AGTGGGTCAG AAGCTTCTTT TTTTGACCCG
1501 GGAGACAGAT GCTGGTGAAC TTCCATCTTC CAGAAATTGT CAAACTATTA
1551 AGTGTCTTTC AAATAAAGAG GTTCTTGAGC AGGCCAGAGT GGAAATCCCC
1601 CTGTTTCCCT TCAGCATTGT TCAGTTCTCT TTTAAGGCCT TTTCACCTGT
1651 CCTCACTGAG GAGATGAACA AAAGGATGAG GATCAAGTGG ACAGAGATAT
1701 CAACTGTCTA TGCTGGGCCA TTTAGCAAAA ATTGCAATCT CAGGGCTCTG
1751 AAGAGGCTGT TTAAAAGCTT TGGCCCAGTC CAGTCAATGA CTTTTGTTCT
1801 TGAAACCCGT CAGGTGCAGA GGCCTGTGAC AGAGCTCACG CTTGATTGTG
1851 ACACCCTCGT GAATGAGCTG GAAGGAGATT CTGAAAACCA AGGCTCTATA
1901 TATCTGTCTG GAGTGAGTGA AACCTTCAAA GAACAGCTAT TGCAGGAGCC
1951 CCGCCTCTTT CTTGGCCTGG AAGCTGTGAT CTTGCCTAAA GATCTTAAAA
2001 GTGGAAAGCA GAAAAAATAC TGTTTCCTGA AATTCAAAAG TTTTGGCAGT
2051 GCCCAGCAGG CCCTCAACAT TCTCACAGGC AAGGACTGGA AGCTGAAAGG
2101 CAGGCATGCC CTAACCCCCA GGCACCTCCA TGCCTGGCTC AGAGGCTTAC
2151 CACCTGAATC AACAAGGCTC CCAGGGCTTC GTGTTGTACC TCCCCCCTTT
2201 GAACAGGAGG CCTTGCAGAC TCTGAAACTG GACCACCCGA AGATAGCAGC
2251 CTGGCGCTGG AGCCGGAAGA TTGGAAAGCT CTACAACAGC TTGTGCCCGG
2301 GCACTCTCTG CCTCATCCTG CTGCCAGGAA CCAAGAGCAC TCATGGTTCA
2351 CTCTCTGGTC TAGGACTGAT GGGAATAAAA GAGGAAGAAG AAAGCGCTGG
2401 CCCAGGCCTG TGTTCGTGAG TCGGCCTGCC ATGTTTCCAT GTGCCATTTC
```

```
2451 TTACCCCTTG TAGGCAATGG CAAAGAATGT GGTCAGGCTG TAGCCTCCCC
2501 AACCAGCAGA CAGTTTTATG GAAACTTGGT ATAGCAGCTA AAAGAGTTTA
2551 GTTTGTTTAT ATGGCATGTA TAAGTTTTCA ATAAATGCCT AAAGTTCAAG
2601 CATAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2651 AGGGCGGCCG CTCTAAAGGA TCCAAGCTTA CGTACGCGAA AAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 188 bp to 2416 bp; peptide length: 743
Category: similarity to known protein

```
  1 MEPEREGTER HPRKVRESRQ APNKLVGAAE AMKAGWDLEE SQPEAKKARL
 51 STILFTDNCE VTHDQLCELL KYAVLGKSNV PKPSWCQLFH QNHLNNVVVF
101 VLQGMSQLHF YRFYLEFGCL RKAFRHKFRL PPPSSDFLAD VVGLQTEQRA
151 GDLPKTMEGP LPSNAKAAIN LQDDPIIQKY GSKKVGLTRC LLTKEEMRTF
201 HFPLQGFPDC ENFLLTKCNG SIADNSPLFG LDCEMCLTSK GRELTRISLV
251 AEGGCCVMDE LVKPENKILD YLTSFSGITK KILNPVTTKL KDVQRQLKAL
301 LPPDAVLVGH SLDLDLRALK MIHPYVIDTS LLYVREQGRR FKLKFLAKVI
351 LGKDIQCPDR LGHDATEDAR TILELARYFL KHGPKKIAEL NLEALANHQE
401 IQAAGQEPKN TAEVLQHPNT SVLECLDSVG QKLLFLTRET DAGELPSSRN
451 CQTIKCLSNK EVLEQARVEI PLFPFSIVQF SFKAFSPVLT EEMNKRMRIK
501 WTEISTVYAG PFSKNCNLRA LKRLFKSFGP VQSMTFVLET RQVQRPVTEL
551 TLDCDTLVNE LEGDSENQGS IYLSGVSETF KEQLLQEPRL FLGLEAVILP
601 KDLKSGKQKK YCFLKFKSFG SAQQALNILT GKDWKLKGRH ALTPRHLHAW
651 LRGLPPESTR LPGLRVVPPP FEQEALQTLK LDHPKIAAWR WSRKIGKLYN
701 SLCPGTLCLI LLPGTKSTHG SLSGLGLMGI KEEEESAGPG LCS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15j3, frame 2

TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product"; Homo sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence., N = 2, Score = 1827, P = 2.1e-284

TREMBL:AF016430_4 gene: "C05C8.5"; Caenorhabditis elegans cosmid C05C8., N = 2, Score = 370, P = 1.7e-34

PIR:S64609 hypothetical protein YGR276c - yeast (Saccharomyces cerevisiae), N = 2, Score = 334, P = 1.8e-27

TREMBLNEW:SPAC637_9 gene: "SPAC637.09"; product: "putative exonuclease"; S.pombe chromosome I cosmid c637., N = 3, Score = 326, P = 2.8e-27

```
>TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product"; Homo
     sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.
          Length = 547

  HSPs:

 Score = 1827 (274.1 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284
 Identities = 358/373 (95%), Positives = 358/373 (95%)

Query:   105 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVVGLQTEQRAGDLPKTMEGPLPSN 164
             MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVVGLQTEQRAGDLPKTMEGPLPSN
Sbjct:     1 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVVGLQTEQRAGDLPKTMEGPLPSN 60

Query:   165 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPLQGFPDCENFLLTKCNGSIAD 224
             AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPLQGFPDCENFLLTKCNGSIAD
Sbjct:    61 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPLQGFPDCENFLLTKCNGSIAD 120
```

```
Query:   225 NSPLFGLDCEM---------------CLTSKGRELTRISLVAEGGCCVMDELVKPENKIL 269
             NSPLFGLDCEM               CLTSKGRELTRISLVAEGGCCVMDELVKPENKIL
Sbjct:   121 NSPLFGLDCEMARTTFNFSIGVLQAECLTSKGRELTRISLVAEGGCCVMDELVKPENKIL 180

Query:   270 DYLTSFSGITKKILNPVTTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDT 329
             DYLTSFSGITKKILNPVTTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDT
Sbjct:   181 DYLTSFSGITKKILNPVTTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDT 240

Query:   330 SLLYVREQGRRFKLKFLAKVILGKDIQCPDRLGHDATEDARTILELARYFLKHGPKKIAE 389
             SLLYVREQGRRFKLKFLAKVILGKDIQCPDRLGHDATEDARTILELARYFLKHGPKKIAE
Sbjct:   241 SLLYVREQGRRFKLKFLAKVILGKDIQCPDRLGHDATEDARTILELARYFLKHGPKKIAE 300

Query:   390 LNLEALANHQEIQAAGQEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDAGELPSSR 449
             LNLEALANHQEIQAAGQEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDAGELPSSR
Sbjct:   301 LNLEALANHQEIQAAGQEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDAGELPSSR 360

Query:   450 NCQTIKCLSNKEV 462
             NCQTIKCLSNKEV
Sbjct:   361 NCQTIKCLSNKEV 373

 Score = 929 (139.4 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284
 Identities = 175/179 (97%), Positives = 177/179 (98%)

Query:   538 LETRQVQRPVTELTLDCDTLVNELEGDSENQGSIYLSGVSETFKEQLLQEPRLFLGLEAV 597
             L  ++VQRPVTELTLDCDTLVNELEGDSENQGSIYLSGVSETFKEQLLQEPRLFLGLEAV
Sbjct:   368 LSNKEVQRPVTELTLDCDTLVNELEGDSENQGSIYLSGVSETFKEQLLQEPRLFLGLEAV 427

Query:   598 ILPKDLKSGKQKKYCFLKFKSFGSAQQALNILTGKDWKLKGRHALTPRHLHAWLRGLPPE 657
             ILPKDLKSGKQKKYCFLKFKSFGSAQQALNILTGKDWKLKGRHALTPRHLHAWLRGLPPE
Sbjct:   428 ILPKDLKSGKQKKYCFLKFKSFGSAQQALNILTGKDWKLKGRHALTPRHLHAWLRGLPPE 487

Query:   658 STRLPGLRVVPPPFEQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK 716
             STRLPGLRVVPPPFEQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK
Sbjct:   488 STRLPGLRVVPPPFEQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTK 546
```

Pedant information for DKFZphtes3_15j3, frame 2

Report for DKFZphtes3_15j3.2

[LENGTH] 743
[MW] 83536.58
[pI] 8.87
[HOMOL] TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product"; Homo sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence. 0.0
[FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 4e-30
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR107w] 3e-13
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YGL094c] 1e-10
[FUNCAT] 04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S. cerevisiae, YGL094c] 1e-10
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-10
[PROSITE] MYRISTYL 5
[PROSITE] AMIDATION 1
[PROSITE] CK2_PHOSPHO_SITE 8
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] GLYCOSAMINOGLYCAN 1
[PROSITE] PKC_PHOSPHO_SITE 16
[PROSITE] ASN_GLYCOSYLATION 2
[PFAM] RNA recognition motif. (aka RRM, RBD, or RNP domain)
[KW] Alpha_Beta

```
SEQ   MEPEREGTERHPRKVRESRQAPNKLVGAAEAMKAGWDLEESQPEAKKARLSTILFTDNCE
PRD   ccchhhhccccchhhhhhhhcchhhhhhhhhhccccccccccchhhhhhccccccccce

SEQ   VTHDQLCELLKYAVLGKSNVPKPSWCQLFHQNHLNNVVVFVLQGMSQLHFYRFYLEFGCL
PRD   eehhhhhhhhhhhhhcccccccccceeeecccccceeeeeeecchhhhhhhhhhhhhh

SEQ   RKAFRHKFRLPPPSSDFLADVVGLQTEQRAGDLPKTMEGPLPSNAKAAINLQDDPIIQKY
PRD   hhhhhhhhcccccccchhhhhhhhhhhhccccccccccccccchhhhhhhhcccccccc

SEQ   GSKKVGLTRCLLTKEEMRTFHFPLQGFPDCENFLLTKCNGSIADNSPLFGLDCEMCLTSK
PRD   ccccccchhhhhhhhhhhhhccccccccccceeeecccccccccccceeeeccccccccc

SEQ   GRELTRISLVAEGGCCVMDELVKPENKILDYLTSFSGITKKILNPVTTKLKDVQRQLKAL
PRD   cchhhhheeeeccccceeeeeeeccccceeeccccccccccccccccchhhhhhhhhhh
```

```
SEQ    LPPDAVLVGHSLDLDLRALKMIHPYVIDTSLLYVREQGRRFKLKFLAKVILGKDIQCPDR
PRD    hccceeeeccccchhhhhhhhhhhhcccccceeeeccccccchhhhhhhhhhhhhhhcccccc

SEQ    LGHDATEDARTILELARYFLKHGPKKIAELNLEALANHQEIQAAGQEPKNTAEVLQHPNT
PRD    cccccchhhhhhhhhhhhhhhhccccceeeeehhhhhhhhhhhhhhhccccccceeeeecccc

SEQ    SVLECLDSVGQKLLFLTRETDAGELPSSRNCQTIKCLSNKEVLEQARVEIPLFPFSIVQF
PRD    ceeeeeeccccceeeeeeccccccccccccccceeeeecchhhhhhhhhhccccccceeee

SEQ    SFKAFSPVLTEEMNKRMRIKWTEISTVYAGPFSKNCNLRALKRLFKSFGPVQSMTFVLET
PRD    eeeceeeehhhhhhhhhhhhheeeeeeccccccchhhhhhhhhhhcccccceeeehhhhh

SEQ    RQVQRPVTELTLDCDTLVNELEGDSENQGSIYLSGVSETFKEQLLQEPRLFLGLEAVILP
PRD    ccccccccccccchhhhhhcccccccccccccccchhhhhhhhhhhhhccccccceeeeec

SEQ    KDLKSGKQKKYCFLKFKSFGSAQQALNILTGKDWKLKGRHALTPRHLHAWLRGLPPESTR
PRD    cccccccccceeeeeeeecccchhhhhhhhhccccccccccccccchhhhhhcccccccc

SEQ    LPGLRVVPPPFEQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTKSTHG
PRD    cccccccccchhhhhhhhhhcchhhhhhhhhhhhhheeeecccccceeeeeccccccccc

SEQ    SLSGLGLMGIKEEEESAGPGLCS
PRD    cccccccchhhhhhccccccccc
```

Prosite for DKFZphtes3_15j3.2

PS00001	219->223	ASN_GLYCOSYLATION	PDOC00001
PS00001	419->423	ASN_GLYCOSYLATION	PDOC00001
PS00002	723->727	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	8->11	PKC_PHOSPHO_SITE	PDOC00005
PS00005	182->185	PKC_PHOSPHO_SITE	PDOC00005
PS00005	238->241	PKC_PHOSPHO_SITE	PDOC00005
PS00005	279->282	PKC_PHOSPHO_SITE	PDOC00005
PS00005	287->290	PKC_PHOSPHO_SITE	PDOC00005
PS00005	447->450	PKC_PHOSPHO_SITE	PDOC00005
PS00005	453->456	PKC_PHOSPHO_SITE	PDOC00005
PS00005	458->461	PKC_PHOSPHO_SITE	PDOC00005
PS00005	481->484	PKC_PHOSPHO_SITE	PDOC00005
PS00005	579->582	PKC_PHOSPHO_SITE	PDOC00005
PS00005	605->608	PKC_PHOSPHO_SITE	PDOC00005
PS00005	630->633	PKC_PHOSPHO_SITE	PDOC00005
PS00005	643->646	PKC_PHOSPHO_SITE	PDOC00005
PS00005	658->661	PKC_PHOSPHO_SITE	PDOC00005
PS00005	678->681	PKC_PHOSPHO_SITE	PDOC00005
PS00005	692->695	PKC_PHOSPHO_SITE	PDOC00005
PS00006	41->45	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006
PS00006	421->425	CK2_PHOSPHO_SITE	PDOC00006
PS00006	458->462	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	630->634	CK2_PHOSPHO_SITE	PDOC00006
PS00007	370->379	TYR_PHOSPHO_SITE	PDOC00007
PS00008	27->33	MYRISTYL	PDOC00008
PS00008	186->192	MYRISTYL	PDOC00008
PS00008	575->581	MYRISTYL	PDOC00008
PS00008	714->720	MYRISTYL	PDOC00008
PS00008	720->726	MYRISTYL	PDOC00008
PS00009	337->341	AMIDATION	PDOC00009

Pfam for DKFZphtes3_15j3.2

```
HMM_NAME     RNA recognition motif. (aka RRM, RBD, or RNP domain)

HMM               *IYVGNLPWDtTEEDLrDlFsQFGpIvsIrMMrDReTGRSRGFAFVEFED
                   IY+ +++ +T +E+L +   + F + + + +++D   G+ + ++F +F++
Query        571   IYLSGVS-ETFKEQLLQEPRLFLGLEAVILPKDLKSGKQKKYCFLKFKS   618

HMM               EEDAekAIdeMNG..meFmGRrIRV*
                   +A+ A+   + G   ++ GR   +
Query        619  FGSAQQALNILTGKDWKLKGRHALT    643
```

DKFZphtes3_15k11
----

group: signal transduction

DKFZphtes3_15k11 encodes a novel 958 amino acid protein C-terminal identical with human KIAA0781 protein and high similarity to protein kinases.

The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The related murine kinase was cloned from the myocardium of the developing heart.

The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

KIAA0781, 5' extension

complete cDNA, complete cds, potential start at Bp 97, EST hits

Sequenced by GBF

Locus: /map="11"

Insert length: 4868 bp
Poly A stretch at pos. 4798, polyadenylation signal at pos. 4776

```
   1 GAGCAAGCGG AGCGGCCGTC GCCCAAGCCA AGCCGCGCTG CCAACCCTCC
  51 CGCCCGCCCG CGCTCCTGTC CGCCGTGTCT AGCAGCGGGG CCCAGCATGG
 101 TCATGGCGGA TGGCCCGAGG CACTTGCAGC GCGGGCCGGT CCGGGTGGGG
 151 TTCTACGACA TCGAGGGCAC GCTGGGCAAG GGCAACTTCG CTGTGGTGAA
 201 GCTGGGGCGG CACCGGATCA CCAAGACGGA GGTGGCAATA AAAATAATCG
 251 ATAAGTCTCA GCTGGATGCA GTGAACCTTG AGAAAATCTA CCGAGAAGTA
 301 CAAATAATGA AAATGTTAGA CCACCCTCAC ATAATCAAAC TTTATCAGGT
 351 AATGGAGACC AAAAGTATGT TGTACCTTGT GACAGAATAT GCCAAAAATG
 401 GAGAAATTTT TGACTATCTT GCTAATCATG GCCGGTTAAA TGAGTCTGAA
 451 GCCAGGCGAA AATTCTGGCA AATCCTGTCT GCTGTTGATT ATTGTCATGG
 501 TCGGAAGATT GTGCACCGTG ACCTCAAAGC TGAAAATCTC CTGCTGGATA
 551 ACAACATGAA TATCAAAATA GCAGATTTCG GTTTTGGAAA TTTCTTTAAA
 601 AGTGGTGAAC TGCTGGCAAC ATGGTGTGGC AGCCCCCCTT ATGCAGCCCC
 651 AGAAGTCTTT GAAGGGCAGC AGTATGAAGG ACCACAGCTG GACATCTGGA
 701 GTATGGGAGT TGTTCTTTAT GTCCTTGTCT GTGGAGCTCT GCCCTTTGAT
 751 GGACCGACTC TTCCAATTTT GAGGCAGAGG GTTCTGGAAG GAAGATTCCG
 801 GATTCCGTAT TTCATGTCAG AAGATTGCGA GCACCTTATC CGAAGGATGT
 851 TGGTCCTAGA CCCATCCAAA CGGCTAACCA TAGCCCAAAT CAAGGAGCAT
 901 AAATGGATGC TCATAGAAGT TCCTGTCCAG AGACCTGTTC TCTATCCACA
 951 AGAGCAAGAA AATGAGCCAT CCATCGGGGA GTTTAATGAG CAGGTTCTGC
1001 GACTGATGCA CAGCCTTGGA ATAGATCAGC AGAAAACCAT TGAGTCTTTG
1051 CAGAACAAGA GCTATAACCA CTTTGCTGCC ATTTATTTCT TGTTGGTGGA
1101 GCGCCTGAAA TCACATCGGA GCAGTTTCCC AGTGGAGCAG AGACTTGATG
1151 GCCGCCAGCG TCGGCCTAGC ACCATTGCTG AGCAAACAGT TGCCAAGGCA
1201 CAGACTGTGG GGCTCCCAGT GACCATGCAT TCACCGAACA TGAGGCTGCT
1251 GCGATCTGCC CTCCTCCCCC AGGCATCCAA CGTGGAGGCC TTTTCATTTC
1301 CAGCATCTGG CTGTCAGGCG GAAGCTGCAT TCATGGAAGA AGAGTGTGTG
1351 GACACTCCAA AGGTCAATGG CTGTCTGCTT GACCCTGTGC CTCCTGTCCT
1401 GGTGCGGAAG GGATGCCAGT CACTGCCCAG CAACATGATG GAGACCTCCA
1451 TTGACGAAGG GCTGGAGACA GAAGGAGAGG CCGAGGAAGA CCCCGCTCAT
1501 GCCTTTGAGG CATTTCAGTC CACACGCAGC GGGCAGAGAC GGCACACTCT
1551 GTCAGAAGTG ACCAATCAAC TGGTCGTGAT GCCTGGGGCA GGGAAAATTT
1601 TCTCCATGAA TGACAGCCCC TCCCTTGACA GTGTGGACTC TGAGTATGAT
1651 ATGGGGTCTG TTCAGAGGGA CCTGAACTTT CTGGAAGACA ACCCTTCCCT
1701 TAAGGACATC ATGTTAGCCA ATCAGCCTTC ACCCCGCATG ACATCTCCCT
1751 TCATAAGCCT GAGACCTACC AACCCAGCCA TGCAGGCTCT GAGCTCCCAG
1801 AAACGAGAGG TCCACAACAG GTCTCCAGTG AGCTTCAGAG AGGGCCGCAG
1851 AGCATCAGAT ACCTCCCTCA CCCAGGGAAT TGTAGCATTT AGACAACATC
1901 TTCAGAATCT GGCTAGAACC AAAGGAATTC TAGAGTTGAA CAAAGTGCAG
1951 TTGTTGTATG AACAAATAGG ACCGGAGGCA GACCCTAACC TGGCGCCGGC
2001 GGCTCCTCAG CTCCAGGACC TTGCTAGCAG CTGCCCTCAG GAAGAAGTTT
2051 CTCAGCAGCA GGAAAGCGTC TCCACTCTCC CTGCCAGCGT GCATCCCCAG
2101 CTGTCCCCAC GGCAGAGCCT GGAGACCCAG TACCTGCAGC ACAGACTCCA
2151 GAAGCCCAGC CTTCTGTCAA AGGCCCAGAA CACCTGTCAG CTTTATTGCA
2201 AAGAACCACC GCGGAGCCTT GAGCAGCAGC TGCAGGAACA TAGGCTCCAG
2251 CAGAAGCGAC TCTTTCTTCA GAAGCAGTCT CAACTGCAGG CCTATTTTAA
2301 TCAGATGCAG ATAGCAGAGA GCTCCTACCC ACAGCCAAGT CAGCAGCTGC
2351 CCCTTCCCCG CCAGGAGACT CCACCGCCTT CTCAGCAGGC CCCACCGTTC
2401 AGCCTGACCC AGCCCCTGAG CCCCGTCCTG GAGCCTTCCT CCGAGCAGAT
2451 GCAATACAGC CCTTTCCTCA GCCAGTACCA AGAGATGCAG CTTCAGCCCC
2501 TGCCCTCCAC TTCCGGTCCC CGGGCTGCTC CTCCTCTGCC CACGCAGCTA
2551 CAGCAGCAGC AGCCGCCACC GCCACCACCC CCTCCACCAC CACGACAGCC
2601 AGGAGCTGCC CCAGCCCCCT TACAGTTCTC CTATCAGACT TGTGAGCTGC
```

```
2651 CAAGCGCTGC TTCCCCTGCG CCAGACTATC CCACTCCCTG TCAGTATCCT
2701 GTGGATGGAG CCCAGCAGAG CGACCTAACG GGGCCAGACT GTCCCAGAAG
2751 CCCAGGACTG CAAGAGGCCC CCTCCAGCTA CGACCCACTA GCCCTCTCTG
2801 AGCTACCTGG ACTCTTTGAT TGTGAAATGC TAGACGCTGT GGATCCACAA
2851 CACAACGGGT ATGTCCTGGT GAATTAGTCT CAGCACAGGA ATTGAGGTGG
2901 GTCAGGTGAA GGAAGAGTGT ATGTTCCTAT TTTTATTCCA GCCTTTTAAA
2951 TTTAAAGCTT ATTTTCTTGC CCTCTCCCTA ACGGGGAGAA ATCGAGCCAC
3001 CCAACTGGAA TCAGAGGGTC TGGCTGGGGT GGATGTTGCT TCCTCCTGGT
3051 TCTGCCCCAC CACAAAGTTT TCTGTGGCAA GTGCTGGAAC ATAGTTGTAG
3101 GCTGAGGCTC CTGCCCTTCG GTCGAGTGGA GCAAGCTCTC GAGGGCAGCA
3151 CTGACAAATG TGTTCCTAAG AAGACATTCA GACCCAGGTC TTATGCAGGA
3201 TTACATCCGT TTATTATCAA GGGCAACCTT GGTGAAAGCA GAAAGGGTGT
3251 GTGCTATTGC ATATATATGG GGGAAAAGGC AATATATTTT TCACTGAAGC
3301 TGAGCAACCA CATATTGCTA CAAGGCAAAT CAAGAAGACA TCAGGAAATC
3351 AGATGCACAG GAAATAAAGG AAAGCTGTGC TTTGTCATTG AATCCTAAGT
3401 TCTTAGCTGC TGATGCAAGT TGTCCCCCAA GGCCATCACA AAGCAGTGGG
3451 GCATGAGCTG TGTTTCAGGG GCCACTAAAT AACAGCTGGT ACTGACCCCA
3501 GAAACCGCCT TCATCTCCAT TCGGAAGCAG GTGACACACC CCTTCAGAAG
3551 GTGCCCTGGG TTGCCGAGTG TCAGAATATA CTCAGGACTC CAGAGGTGTC
3601 ACACGTGGAA CTGACAGGAG ACCCGCCACC GTGGAGGCAG GGGGCAAGAA
3651 ACTCAAGAAC GCATCAAGAG CACCAGCCCT GGGCCAGGGA AGACAGGCTC
3701 TTCCTGCAGT TTCTCGTGGA CACTGCTGGC TTGCGGGCAG TCGGTCTCCA
3751 GGGTACCTGT TGTCTCTTTT CCGATGTAAT AACTACTTTG ACCTTACACT
3801 ATATGTTGCT AGTAGTTTAT TGAGCTTTGT ATATTTGGAC AGTTTCATAT
3851 AGGGCTTAGA GATTTTAAGG ACATGATAAA TGAACTTTTC TGTCCCATGT
3901 GAAGTGGTAG TGCGGTGCCT TTCCCCCAGA TCATGCTTTA ATTCTTTCTT
3951 TTCTGTAGAA ACCAACAGTT TCCATTTATG TCAATGCTAA ATCCAAAGTC
4001 ACTTCAGAGT TTGTTTTCCA CCATGTGGGA ATCAGCATTC TTAATTTCGT
4051 TAAAGTTTTG ACTTGTAATG AAATGTTCAA GTATTACAGC AATATTCAAA
4101 GAAAGAACCA CAGATGTGTT AACCATTTAA GCAGATCATC TGCCAAACAT
4151 TATATTACTA ATAAAACTTA ACCAACACTT ACAATTCAGT CATCAAAGTA
4201 AGTAAAAATT AGATGCTACA GCTAGCTAAC TGTATCCCTA GAAATGATGA
4251 ATAATTTGCC ATTTGGACAG TTAACATCCA GGTGTTACAA AGTCAGTGTT
4301 AATTCTAAAG ATGATCATTT CTGCCCTTTA GAATGGCTTG TCCCATCAGC
4351 AGATGAATGT GTTAAGCACA AAGCATCTTC CTTAAAGCAC AAAGAGAGGG
4401 ACTAACTGAT GCTGCATCTA GAAAACACCT TTAAGTTGCC TTTCCTCTTT
4451 GTAGTTAGCG TTCAGGCAGG TGACGTGTGG AAAGTCTAGG GGGTTCCATT
4501 CTGGCCATGC GAGCCCAGCT CCTACCAACG TCGGTAACTT GAGCAGTCCC
4551 TGTTGCTGGC CAGAGACTGC CTGGTCGCCA GCGCTCACCA TGGGTGCCAG
4601 GATGCTTCGC AGAGGCACTG TGCTCACGGT TGGACTTGGT GTCAGTGGGA
4651 AAGGGCAGTG TGGGGACTGT CATTTTTGTG ATTTAATAAC ACACAGTGAA
4701 AATCCAGGAA GAATGAATTA AGCTTCTTCT GGGAGTTGTT TATTCCTGCT
4751 CGTGCTTAAG ATTGATGATT TCGTGAAATA AAGAACATCA TTTCATTTAA
4801 AAAAAAAAAA AAAAAAAGGG CGGCCGCTCT AGAGGATCCA AGCTTACGTA
4851 CGCGTGAAAA AAAAAAAG
```

BLAST Results

Entry HSG4921 from database EMBL:
human STS SHGC-37164.
Score = 1605, P = 1.9e-66, identities = 349/369

Entry AB018324 from database EMBL:
Homo sapiens mRNA for KIAA0781 protein, partial cds.
Score = 10725, P = 0.0e+00, identities = 2145/2145

Medline entries

No Medline entry

Peptide information for frame 1

ORF from the beginning to 2874 bp; peptide length: 959
Category: known protein

```
  1 EQAERPSPKP SRAANPPARP RSCPPCLAAG PSMVMADGPR HLQRGPVRVG
 51 FYDIEGTLGK GNFAVVKLGR HRITKTEVAI KIIDKSQLDA VNLEKIYREV
101 QIMKMLDHPH IIKLYQVMET KSMLYLVTEY AKNGEIFDYL ANHGRLNESE
151 ARRKFWQILS AVDYCHGRKI VHRDLKAENL LLDNNMNIKI ADFGFGNFFK
201 SGELLATWCG SPPYAAPEVF EGQQYEGPQL DIWSMGVVLY VLVCGALPFD
251 GPTLPILRQR VLEGRFRIPY FMSEDCEHLI RRMLVLDPSK RLTIAQIKEH
```

```
301 KWMLIEVPVQ RPVLYPQEQE NEPSIGEFNE QVLRLMHSLG IDQQKTIESL
351 QNKSYNHFAA IYFLLVERLK SHRSSFPVEQ RLDGRQRRPS TIAEQTVAKA
401 QTVGLPVTMH SPNMRLLRSA LLPQASNVEA FSFPASGCQA EAAFMEEECV
451 DTPKVNGCLL DPVPPVLVRK GCQSLPSNMM ETSIDEGLET EGEAEEDPAH
501 AFEAFQSTRS GQRRHTLSEV TNQLVVMPGA GKIFSMNDSP SLDSVDSEYD
551 MGSVQRDLNF LEDNPSLKDI MLANQPSPRM TSPFISLRPT NPAMQALSSQ
601 KREVHNRSPV SFREGRRASD TSLTQGIVAF RQHLQNLART KGILELNKVQ
651 LLYEQIGPEA DPNLAPAAPQ LQDLASSCPQ EEVSQQQESV STLPASVHPQ
701 LSPRQSLETQ YLQHRLQKPS LLSKAQNTCQ LYCKEPPRSL EQQLQEHRLQ
751 QKRLFLQKQS QLQAYFNQMQ IAESSYPQPS QQLPLPRQET PPPSQQAPPF
801 SLTQPLSPVL EPSSEQMQYS PFLSQYQEMQ LQPLPSTSGP RAAPPLPTQL
851 QQQQPPPPPP PPPPRQPGAA PAPLQFSYQT CELPSAASPA PDYPTPCQYP
901 VDGAQQSDLT GPDCPRSPGL QEAPSSYDPL ALSELPGLFD CEMLDAVDPQ
951 HNGYVLVN
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_15k11, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_15k11, frame 1

Report for DKFZphtes3_15k11.1

[LENGTH] 926
[MW] 103915.77
[pI] 5.70
[HOMOL] TREMBL:AB018324_1 gene: "KIAA0781"; product: "KIAA0781 protein"; Homo sapiens mRNA for KIAA0781 protein, partial cds. 0.0
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w] 8e-76
[FUNCAT] 11.01 stress response [S. cerevisiae, YDR477w] 8e-76
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 8e-76
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YCL024w] 4e-58
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 3e-56
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c] 3e-56
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 1e-53
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 3e-53
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL101w] 3e-53
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 5e-51
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 5e-42
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 2e-34
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-26
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YER129w] 3e-26
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae, YPL031c] 1e-23
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c] 1e-23
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 1e-23
[FUNCAT] 03.13 meiosis [S. cerevisiae, YOR351c] 2e-23
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YHL007c] 8e-21
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YHL007c] 8e-21
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YPL140c] 2e-20
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YLR113w] 7e-20
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w] 3e-19
[FUNCAT] 10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 2e-18
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YLR362w] 3e-18
[FUNCAT] 04.03.99 other trna-transcription activities [S. cerevisiae, YOR061w] 4e-18
[FUNCAT] 06.07 protein modification (glycolsylation, acylation, myristylation, palmitylation, farnesylation and processing) [S. cerevisiae, YFL033c] 4e-17
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 2e-16
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae, YNL183c] 2e-14

[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c] 2e-14
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YNL020c] 5e-14
[FUNCAT] c energy conversion [M. genitalium, MG109] 2e-12
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YBR097w] 1e-10
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w] 1e-10
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YBR097w] 1e-10
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w] 1e-10
[FUNCAT] 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w] 4e-09
[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YHR079c] 1e-07
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c] 1e-07
[FUNCAT] 08.19 cellular import [S. cerevisiae, YNL154c] 2e-04
[BLOCKS] BL00415A Synapsins proteins
[BLOCKS] BL00239B Receptor tyrosine kinase class II proteins
[BLOCKS] BL00107A Protein kinases ATP-binding region proteins
[SCOP] d1gol__ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus 3e-78
[SCOP] d1wfc__ 5.1.1.1.8 MAP kinase p38 (human (Homo sapiens) 1e-81
[SCOP] d1koa_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditi 5e-89
[SCOP] d1koba_ 5.1.1.1.6 Twitchin, kinase domain (california sea har 5e-86
[SCOP] d1phk__ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinas 3e-80
[SCOP] d1irk__ 5.1.1.2.4 insulin receptor [Human (Homo sapiens) 6e-70
[SCOP] d1apme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu 1e-95
[SCOP] d1fgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Hom 7e-71
[SCOP] d1ydse_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo 2e-96
[SCOP] d1fmk_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Hom 2e-72
[SCOP] d1cdka_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su 5e-97
[SCOP] d2hckb3 5.1.1.2.1 (167-437) Haemopoetic cell kinase Hck [huma 2e-68
[SCOP] d1csn__ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe 3e-53
[SCOP] d1jsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens) 3e-78
[SCOP] d1ckia_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus) 1e-58
[EC] 2.7.1.117 Myosin-light-chain kinase 3e-49
[EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 4e-78
[EC] 2.7.1.38 Phosphorylase kinase 3e-41
[EC] 2.7.1.37 Protein kinase 7e-45
[EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 5e-42
[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 4e-78
[PIRKW] phosphotransferase 3e-93
[PIRKW] nucleus 2e-74
[PIRKW] calcium 2e-40
[PIRKW] transferase 3e-33
[PIRKW] duplication 2e-32
[PIRKW] tandem repeat 7e-45
[PIRKW] phorbol ester binding 4e-33
[PIRKW] zinc 4e-33
[PIRKW] ion transport 1e-32
[PIRKW] cell cycle control 1e-45
[PIRKW] serine/threonine-specific protein kinase 2e-97
[PIRKW] oncogene 1e-34
[PIRKW] phospholipid binding 2e-32
[PIRKW] autophosphorylation 2e-74
[PIRKW] brain 6e-36
[PIRKW] heterotetramer 8e-38
[PIRKW] mitosis 1e-45
[PIRKW] polymer 5e-41
[PIRKW] magnesium 6e-80
[PIRKW] ATP 2e-97
[PIRKW] polyprotein 1e-34
[PIRKW] alternative initiators 2e-31
[PIRKW] phosphoprotein 2e-74
[PIRKW] apoptosis 8e-38
[PIRKW] cGMP binding 4e-33
[PIRKW] glycoprotein 3e-36
[PIRKW] skeletal muscle 8e-38
[PIRKW] protein kinase 2e-50
[PIRKW] testis 5e-41
[PIRKW] cAMP binding 8e-38
[PIRKW] transforming protein 4e-33
[PIRKW] purine nucleotide binding 7e-52
[PIRKW] calcium binding 7e-45
[PIRKW] alternative splicing 5e-42
[PIRKW] P-loop 7e-52
[PIRKW] lipoprotein 8e-38
[PIRKW] proto-oncogene 4e-33
[PIRKW] segmentation 1e-34
[PIRKW] core protein 1e-34

[PIRKW] muscle 8e-38
[PIRKW] myristylation 8e-38
[PIRKW] EF hand 7e-45
[PIRKW] cell division 3e-49
[PIRKW] homodimer 1e-32
[PIRKW] calmodulin binding 5e-42
[SUPFAM] ribosomal protein S6 kinase II 1e-34
[SUPFAM] calcium-dependent protein kinase 7e-45
[SUPFAM] AMP-activated protein kinase 6e-80
[SUPFAM] protein kinase akt 3e-36
[SUPFAM] protein kinase SPK1 7e-41
[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 8e-99
[SUPFAM] Ca2+/calmodulin-dependent protein kinase 5e-42
[SUPFAM] calmodulin repeat homology 7e-45
[SUPFAM] cAMP receptor protein cyclic nucleotide-binding domain homology 3e-33
[SUPFAM] protein kinase DUN1 6e-36
[SUPFAM] protein kinase C zeta 4e-33
[SUPFAM] Dictyostelium cAMP-dependent protein kinase catalytic chain 2e-34
[SUPFAM] death-associated protein kinase 8e-38
[SUPFAM] pleckstrin repeat homology 3e-36
[SUPFAM] ankyrin repeat homology 8e-38
[SUPFAM] protein kinase homology 8e-99
[SUPFAM] Ca2+/calmodulin-dependent protein kinase II 6e-38
[SUPFAM] protein kinase C zinc-binding repeat homology 4e-33
[SUPFAM] protein kinase C delta 2e-32
[SUPFAM] cGMP-dependent protein kinase 3e-33
[SUPFAM] protein kinase cdr1 1e-45
[SUPFAM] kinase-related transforming protein 2e-50
[SUPFAM] Ca2+/calmodulin-dependent protein kinase I 8e-42
[SUPFAM] kinase interaction domain homology 7e-41
[SUPFAM] gag-akt polyprotein 1e-34
[PROSITE] PROTEIN_KINASE_ATP 1
[PROSITE] MYRISTYL 3
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 4
[PROSITE] CK2_PHOSPHO_SITE 15
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 10
[PROSITE] ASN_GLYCOSYLATION 2
[PROSITE] PROTEIN_KINASE_ST 1
[PFAM] Eukaryotic protein kinase domain
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 12.31 %

```
SEQ    MVMADGPRHLQRGPVRVGFYDIEGTLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDAVN
SEG    ............................................................
1ctpE  .......................EEECTTTEEEEEEEETTTTEEEEEEEEEHHHHHHHC

SEQ    LEKIYREVQIMKMLDHPHIIKLYQVMETKSMLYLVTEYAKNGEIFDYLANHGRLNESEAR
SEG    ............................................................
1ctpE  HHHHHHHHHHHHCCCTTTBCCEEEEEEEETTEEEEEEECTTTTBHHHHHHHHCCCCHHHHH

SEQ    RKFWQILSAVDYCHGRKIVHRDLKAENLLLDNNMNIKIADFGFGNFFKSGELLATWCGSP
SEG    ............................................................
1ctpE  HHHHHHHHHHHHHHHCCEECCCCCGGGEEETTTTCEEECCTTTTEETT-TTBC-CCCCCG

SEQ    PYAAPEVFEGQQYEGPQLDIWSMGVVLYVLVCGALPFDGPTLPILRQRVLEGRFRIPYFM
SEG    ............................................................
1ctpE  GGCCHHHHHCCCBC-HHHHHHHHHHHHHHHHHCCTTTTTTTHHHHHHHHHHCCCCCTTTT

SEQ    SEDCEHLIRRMLVLDPSKRLTIAQIKEHKWMLIEVPVQRPVLYPQEQENEPSIGEFNEQV
SEG    ............................................................
1ctpE  CHHHHHHHHHTTTTGGGTTTHHHHHHCGG...............................

SEQ    LRLMHSLGIDQQKTIESLQNKSYNHFAAIYFLLVERLKSHRSSFPVEQRLDGRQRRPSTI
SEG    ............................................................
1ctpE  ............................................................

SEQ    AEQTVAKAQTVGLPVTMHSPNMRLLRSALLPQASNVEAFSFPASGCQAEAAFMEEECVDT
SEG    ............................................................
1ctpE  ............................................................

SEQ    PKVNGCLLDPVPPVLVRKGCQSLPSNMMETSIDEGLETEGEAEEDPAHAFEAFQSTRSGQ
SEG    ................................xxxxxxxxxxx.................
1ctpE  ............................................................

SEQ    RRHTLSEVTNQLVVMPGAGKIFSMNDSPSLDSVDSEYDMGSVQRDLNFLEDNPSLKDIML
SEG    ............................................................
1ctpE  ............................................................
```

```
SEQ    ANQPSPRMTSPFISLRPTNPAMQALSSQKREVHNRSPVSFREGRRASDTSLTQGIVAFRQ
SEG    ............................................................
1ctpE  ............................................................

SEQ    HLQNLARTKGILELNKVQLLYEQIGPEADPNLAPAAPQLQDLASSCPQEEVSQQQESVST
SEG    ...........................xxxxxxxxxxxxxxxx....xxxxxxxxxxxx.
1ctpE  ............................................................

SEQ    LPASVHPQLSPRQSLETQYLQHRLQKPSLLSKAQNTCQLYCKEPPRSLEQQLQEHRLQQK
SEG    ...............................................xxxxxxxxxxxxx
1ctpE  ............................................................

SEQ    RLFLQKQSQLQAYFNQMQIAESSYPQPSQQLPLPRQETPPPSQQAPPFSLTQPLSPVLEP
SEG    xxxxxxxxxxx..........xxxxxxxxxxxxxxx........................
1ctpE  ............................................................

SEQ    SSEQMQYSPFLSQYQEMQLQPLPSTSGPRAAPPLPTQLQQQQPPPPPPPPPPRQPGAAPA
SEG    ...........................xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
1ctpE  ............................................................

SEQ    PLQFSYQTCELPSAASPAPDYPTPCQYPVDGAQQSDLTGPDCPRSPGLQEAPSSYDPLAL
SEG    xxx.........................................................
1ctpE  ............................................................

SEQ    SELPGLFDCEMLDAVDPQHNGYVLVN
SEG    ..........................
1ctpE  ..........................
```

Prosite for DKFZphtes3_15k11.1

PS00001	115->119	ASN_GLYCOSYLATION	PDOC00001
PS00001	320->324	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	355->359	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	481->485	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	584->588	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	475->478	PKC_PHOSPHO_SITE	PDOC00005
PS00005	534->537	PKC_PHOSPHO_SITE	PDOC00005
PS00005	545->548	PKC_PHOSPHO_SITE	PDOC00005
PS00005	554->557	PKC_PHOSPHO_SITE	PDOC00005
PS00005	567->570	PKC_PHOSPHO_SITE	PDOC00005
PS00005	579->582	PKC_PHOSPHO_SITE	PDOC00005
PS00005	670->673	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	359->363	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	450->454	CK2_PHOSPHO_SITE	PDOC00006
PS00006	458->462	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	534->538	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	878->882	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	672->680	TYR_PHOSPHO_SITE	PDOC00007
PS00007	100->108	TYR_PHOSPHO_SITE	PDOC00007
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	871->877	MYRISTYL	PDOC00008
PS00008	905->911	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009
PS00009	582->586	AMIDATION	PDOC00009
PS00107	26->50	PROTEIN_KINASE_ATP	PDOC00100
PS00108	138->151	PROTEIN_KINASE_ST	PDOC00100

Pfam for DKFZphtes3_15k11.1

HMM_NAME Eukaryotic protein kinase domain

```
HMM              *YeigRiIGeGsFGtVYkCiWr.TGeIVAIKIIkkrsms......FlREI
                  Y I++++G+G+F++V+++++R T  +VAIKII+K++++      + RE+
Query        20  YDIEGTLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDAVNLEKIYREV    68

HMM              qIMRrLnHPNIIRFYDwFeddddHIYMIMEYMeGGDLFDYIrrngpMsEw
                 QIM++L+HP+II++Y ++E +++ +Y+++EY+ +G++FDY+ ++G+++E
Query        69  QIMKMLDHPHIIKLYQVME-TKSMLYLVTEYAKNGEIFDYLANHGRLNES   117

HMM              eIrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeNgqIKIcDFGLARqM
                 E+R+ ++QIL++++Y+H ++I+HRDLK+EN+L+D+N++IKI+DFG+  ++
Query       118  EARRKFWQILSAVDYCHGRKIVHRDLKAENLLLDNNMNIKIADFGFGNFF   167

HMM              nnYerMttfCGTPWYMMAPEVIImg.nyYttkVDMWSFGCILWEMMTGep
                 +++E++ T CG+P+Y  APEV  +G +Y +++ D+WS+G++L+ +++G +
Query       168  KSGELLATWCGSPPYA-APEV-FEGQQYEGPQLDIWSMGVVLYVLVCGAL   215

HMM              PFyddnMemImrIiqrfrrpfWpnCSeElyDFMrwCWnyDPekRPTFrQI
                 PF++ ++  + + +++ R+++++ +SE++ +++R+++ +DP+KR+T+ QI
Query       216  PFDGPTLPILRQRVLEGRFRIPYFMSEDCEHLIRRMLVLDPSKRLTIAQI   265

HMM              LnHPWF*
                  +H W+
Query       266  KEHKWM    271
```

DKFZphtes3_17f10
----

group: testes derived

DKFZphtes3_15j18 encodes a novel 710 amino acid protein with weak similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

Sequenced by GBF

Locus: unknown

Insert length: 2533 bp
Poly A stretch at pos. 2507, no polyadenylation signal found

```
   1 CTTCAGTTCA ACTAAAAATG GACAGATCTC AGCAGACCAG CCGTACAGGA
  51 TACTGGACCA TGATGAACAT CCCCCCTGTA GAAAAAGTGG ACAAGGAACA
 101 ACAGACATAC TTTAGTGAAT CAGAAATAGT GGTTATTTCC AGGCCAGATA
 151 GTTCTTCTAC AAAGTCAAAG GAAGATGCCC TGAAACATAA ATCGTCGGGA
 201 AAGATTTTTG CTAGTGAACA CCCTGAATTT CAACCAGCAA CAAACAGCAA
 251 TGAAGAAATT GGGCAGAAAA ATATCAGCAG AACTTCATTT ACTCAGGAGA
 301 CTAAAAAAGG TCCCCCAGTA CTTTTAGAAG ATGAGCTTAG GGAAGAAGTA
 351 ACTGTACCTG TTGTACAAGA AGGTTCTGCT GTTAAAAAAG TGGCTTCTGC
 401 TGAAATAGAG CCTCCATCAA CAGAAAAATT CCCAGCTAAA ATACAGCCTC
 451 CATTAGTTGA AGAGGCCACT GCTAAAGCGG AGCCCAGACC TGCTGAAGAG
 501 ACCCATGTCC AAGTACAGCC ATCAACTGAA GAGACTCCTG ATGCTGAGGC
 551 AGCCACTGCA GTTGCGGAGA ATTCTGTTAA AGTTCAGCCT CCACCTGCTG
 601 AAGAGGCCCC TTTAGTGGAG TTTCCTGCTG AAATTCAGCC TCCATCAGCT
 651 GAAGAGTCTC CTTCTGTAGA GCTTCTGGCT GAAATTCTGC CTCCATCAGC
 701 TGAAGAGTCC CCTTCAGAAG AGCCTCCTGC TGAAATTCTG CCTCCACCAG
 751 CTGAAAAGTC TCCTTCAGTA GAGCTTCTTG GTGAAATTCG GTCTCCCTCA
 801 GCACAAAAGG CTCCCATTGA AGTACAGCCT TTACCAGCTG AGGGCGCCCT
 851 TGAAGAGGCC CCAGCTAAAG TAGAGCCTCC CACTGTTGAA GAGACCCTTG
 901 CTGAAGTTCA GCCTCTATTA CCTGAAGAGG CTCCTAGAGA AGAGGCTCGA
 951 GAACTTCAGC TTTCAACAGC TATGGAGACC CCTGCAGAAG AGGCTCCTAC
1001 TGAATTTCAG TCTCCATTAC CTAAAGAGAC CACTGCAGAA GAGGCCTCTG
1051 CTGAAATTCA GCTTCTAGCA GCTACGGAGC CTCCTGCAGA TGAAACTCCT
1101 GCCGAAGCTC GGTCTCCACT ATCTGAGGAG ACTTCTGCAG AAGAGGCTCA
1151 TGCTGAAGTT CAATCTCCAT TAGCTGAAGA GACCACTGCA GAAGAGGCCT
1201 CTGCTGAAAT TCAGCTTCTA GCAGCTATAG AGGCTCCTGC AGATGAAACT
1251 CCTGCTGAAG CTCAGTCTCC ACTATCTGAG GAGACTTCTG CAGAAGAGGC
1301 TCCTGCTGAA GTTCAGTCTC CATCAGCTAA GGGAGTTTCT ATAGAAGAGG
1351 CCCCTCTTGA GCTTCAGCCT CCATCAGGTG AAGAGACCAC TGCAGAAGAG
1401 GCCTCTGCTG CAATTCAGCT TCTAGCAGCT ACAGAGGCTT CTGCAGAAGA
1451 GGCTCCTGCT GAAGTTCAGC CTCCACCAGC TGAGGAGGCC CCCGCTGAAG
1501 TTCAGCCTCC ACCAGCTGAG GAGGCCCCCG CTGAAGTTCA GCCTCCACCA
1551 GCTGAGGAGG CCCCCGCTGA AGTTCAGCCT CCACCAGCTG AGGAGGCCCC
1601 CGCTGAAGTT CAGCCTCCAC CAGCTGAGGA GGCCCCCGCT GAAGTTCAGC
1651 CTCCACCAGC TGAGGAGGCC CCCTCTGAAG TTCAGCCTCC ACCAGCTGAG
1701 GAGGCCCCTG CTGAAGTTCA GTCTCTACCA GCTGAGGAGA CTCCTATAGA
1751 AGAGACCCTT GCTGCAGTAC ACTCTCCCCC AGCTGATGAT GTCCCTGCAG
1801 AAGAGGCCTC CGTTGACAAA CATTCCCCAC CAGCTGATTT GCTTCTGACT
1851 GAGGAGTTTC CTATAGGAGA GGCCTCTGCT GAAGTTTCAC CTCCACCATC
1901 TGAACAAACC CCTGAAGATG AGGCTCTGGT AGAGAATGTG TCTACAGAAT
1951 TTCAGTCACC GCAGGTGGCA GGAATTCCAG CAGTAAAATT AGGATCGGTT
2001 GTTTTGGAAG GTGAAGCAAA ATTTGAAGAG GTTTCAAAAA TCAATTCTGT
2051 CCTTAAAGAT TTGTCTAATA CCAATGATGG ACAGGCTCCC ACTCTTGAAA
2101 TAGAAAGTGT TTTTCATATA GAATTAAAAC AACGTCCTCC TGAACTGTAG
2151 TCAGGTTGTA CCTAAGCTAG CAATCAGAAG CTACATGGTT TTGGAAGAAC
2201 ATACTTTAGA AAAGGGTGGG CAGCAGGAAG TAGCTTTGTC AATAAGGCAA
2251 ATTAAAGGGG ACCCCAAGAC TTGGAATACA GGTTGGAAAA TGAACAATAA
2301 AAACTGTAGC AGCATAAAAT TACTTGTGTT AATTTCATTC AAATTTATGG
2351 CATGAAAAAT ACCTATTTTG AAAGTAAGTT TATAATTGAA AAAAATTGCT
2401 TAAAATATCC TTCCTACAGT AAACTTGTTG ACACGAGTAA AGTTTAATCT
2451 GCAGCCATCT TTTCTTGTCT TTGCCTTCCC TTTATAAGTA AATATAGTTT
2501 CTAGTGGAAA AAAAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 18 bp to 2147 bp; peptide length: 710
Category: similarity to known protein
Classification: unclassified

```
  1 MDRSQQTSRT GYWTMMNIPP VEKVDKEQQT YFSESEIVVI SRPDSSSTKS
 51 KEDALKHKSS GKIFASEHPE FQPATNSNEE IGQKNISRTS FTQETKKGPP
101 VLLEDELREE VTVPVVQEGS AVKKVASAEI EPPSTEKFPA KIQPPLVEEA
151 TAKAEPRPAE ETHVQVQPST EETPDAEAAT AVAENSVKVQ PPPAEEAPLV
201 EFPAEIQPPS AEESPSVELL AEILPPSAEE SPSEEPPAEI LPPPAEKSPS
251 VELLGEIRSP SAQKAPIEVQ PLPAEGALEE APAKVEPPTV EETLAEVQPL
301 LPEEAPREEA RELQLSTAME TPAEEAPTEF QSPLPKETTA EEASAEIQLL
351 AATEPPADET PAEARSPLSE ETSAEEAHAE VQSPLAEETT AEEASAEIQL
401 LAAIEAPADE TPAEAQSPLS EETSAEEAPA EVQSPSAKGV SIEEAPLELQ
451 PPSGEETTAE EASAAIQLLA ATEASAEEAP AEVQPPPAEE APAEVQPPPA
501 EEAPAEVQPP PAEEAPAEVQ PPPAEEAPAE VQPPPAEEAP AEVQPPPAEE
551 APSEVQPPPA EEAPAEVQSL PAEETPIEET LAAVHSPPAD DVPAEEASVD
601 KHSPPADLLL TEEFPIGEAS AEVSPPPSEQ TPEDEALVEN VSTEFQSPQV
651 AGIPAVKLGS VVLEGEAKFE EVSKINSVLK DLSNTNDGQA PTLEIESVFH
701 IELKQRPPEL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17f10, frame 3

PIR:A37221 neurofilament triplet H protein - rat, N = 1, Score = 480, P = 7.4e-43

TREMBL:RNNFLH_1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N = 1, Score = 475, P = 1e-42

```
>PIR:A37221 neurofilament triplet H protein - rat
          Length = 1,072

  HSPs:

 Score = 480 (72.0 bits), Expect = 7.4e-43, P = 7.4e-43
 Identities = 185/622 (29%), Positives = 320/622 (51%)

Query:    33 SESEIVVISRPDSSSTKSKEDALKHKSSGKIFASEHPEFQPATNSNEEIGQKNISRTSFT 92
             SE +I V+ + + +    +E    + + + ++    E  E Q       E G +  + TS
Sbjct:   436 SEEKIKVVEKSEKETVIVEEQTEEIQVTEEVTEEEDKEAQGEEEEEAEEGGEEAATTSPP 495

Query:    93 QETKKGPPVLLEDELREEVTVPVVQEGSAVKKVASAEIEPPSTEKFPAKIQPPLVEEATA 152
               E    P    +  ++EE    P    +  A  K + AE + P+  K PA+++ P   ++ A
Sbjct:   496 AEEAASPEKETKSPVKEEAKSPAEAKSPAEAK-SPAEAKSPAEVKSPAEVKSPAEAKSPA 554

Query:   153 KAEPRPAEETHVQVQPSTEETPDAEAATAVAENSVKVQPPPAEEAP-LVEFPAEIQPPSA 211
             +A+  PAE   V+  P+T ++P     + A A++  +V+ P    ++P    + PAE + P+
Sbjct:   555 EAKS-PAE---VK-SPATVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAE 609

Query:   212 EESP-SVELLAEILPPSAEESPSE-EPPAEILPPPAEKSPS-VELLGEIRSPSAQKAPIE 268
             +SP    +  AE   P++  +SP E + PAE     P   KSP+  V+  E +SP+  K+P+
Sbjct:   610 VKSPVEAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKSPVT 669

Query:   269 VQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLPEEAPREEARELQLSTAMETPAE-EAP 327
             V+   PAE      ++P +V+ P    ++ +E +      ++P E A+          ++PAE ++P
Sbjct:   670 VKS-PAEA---KSPVEVKSPASVKSPSEAKSPAGAKSPAE-AKS---PVVAKSPAEAKSP 721

Query:   328 TEFQSPLPKETTAEEASAEIQLLAATEPPAD-ETPAEARSPLSEETSAEEAHAEVQS--- 383
               E  + P    ++  AE  S          A   + PA+ ++PAEA+SP+ E   S E+A  + V+
Sbjct:   722 AEAKPPAEAKSPAEAKSP-----AEAKSPAEAKSPAEAKSPV-EVKSPEKAKSPVKEGAK 775

Query:   384 PLAEETTAEEASAEIQLLAAIEAPAD-ETPAEAQSPLSEET-SAEEAPA-EVQSPSAKGV 440
```

```
               LAE  + E+A + ++    I+ PA+ ++P +A+SP+ EE  S E+A  +V+SP AK
Sbjct:   776 SLAEAKSPEKAKSPVK--EEIKPPAEVKSPEKAKSPMKEEAKSPEKAKTLDVKSPEAKTP 833

Query:   441 SIEEA--PLELQPPSGEETTA-EEASAAIQLLAATEASA---EEAPAEVQPPPAEEAPAE 494
             + EEA  P +++ P   ++ A EEA +  +     TE  A    EE  + V+   A+E P +
Sbjct:   834 AKEEAKRPADIRSPEQVKSPAKEEAKSPEKEETRTEKVAPKKEEVKSPVEEVKAKEPPKK 893

Query:   495 VQPPPAEEAP-AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPS 553
             V+        P  EV+    +EAP E Q P AEE      + P  +++P E +     EEA
Sbjct:   894 VEEEKTPATPKTEVKESKKDEAPKEAQKPKAEEKEPLTEKP--KDSPGEAKK---EEAKE 948

Query:   554 EVQPPPAEEAPAEV----QSLP---AEETPIEETL--AAVHSPPADDVPAEEASVD-KHS 603
             +     P EE PA++     ++ P    AE+    +E     +     P ++VPA     D K
Sbjct:   949 KKAAAPEEETPAKLGVKEEAKPKEKAEDAKAKEPSKPSEKEKPKKEEVPAAPEKKDTKEE 1008

Query:   604 PPADLLLTEEFPIGEASAEVSPP--PSEQT-PEDEALVENVSTEFQSPQ 649
               +      EE P  +A A+     P E + P+ E     ++ ST+ + Q
Sbjct:  1009 KTTESKKPEEKPKMQAKAKEEDKGLPQEPSKPKTEKAEKSSSTDQKDSQ 1057

 Score = 473 (71.0 bits), Expect = 4.8e-42, P = 4.8e-42
 Identities = 184/628 (29%), Positives = 310/628 (49%)

Query:    18 IPPVEKVDKEQQTYFSESEIVVISRP---DSSSTKSKEDALKHKSSGKIFASEHPEFQPA 74
             I  VEK +KE       ++E + ++      +              E+  + + G+  A+  P  + A
Sbjct:   440 IKVVEKSEKETVIVEEQTEEIQVTEEVTEEEDKEAQGEEEEEAEEGGEEAATTSPPAEEA 499

Query:    75 TNSNEEIGQKNISRTSFTQETKKGPPVLLEDELREEVTVPVVQEGSAVKKVASAEIEPPS 134
             +  +E  +  +  + +   K P  E+  E  P  + A  K + AE + P+
Sbjct:   500 ASPEKET-KSPVKEEAKSPAEAKSPA---EAKSPAEAKSPAEVKSPAEVK-SPAEAKSPA 554

Query:   135 TEKFPAKIQPPLVEEATAKAEPRPAEETHVQVQ-PSTEETPDAEAATAVAENSVKVQPPP 193
             K PA+++ P  ++ A+A+  ++ +V+ P+T ++P   + A A++ +V+ P
Sbjct:   555 EAKSPAEVKSPATVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV 614

Query:   194 AEEAPL-VEFPAEIQPPSAEESPS-VELLAEILPPSAEESPSE-EPPAEILPPPAEKSPS 250
             ++P  + PA ++ P  +SP+  + AE+ P+  +SP E + PAE+ P  KSP+
Sbjct:   615 EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKSPVTVKSPA 674

Query:   251 -VELLGEIRSPSAQKAPIEVQ-PLPAEGALE-EAPAKVEPPTVEETLAEVQPLLPEEAPR 307
             +   E++SP++ K+P E + P A+  E ++P + P ++ AE +P ++P
Sbjct:   675 EAKSPVEVKSPASVKSPSEAKSPAGAKSPAEAKSPVVAKSPAEAKSPAEAKPPAEAKSPA 734

Query:   308 EEARELQLSTAME--TPAE-EAPTEFQSP----LP-KE---TTAEEASAEIQLLAATE-- 354
             E  +  + E  +PAE ++P E +SP  P KE + AE S E E
Sbjct:   735 EAKSPAEAKSPAEAKSPAEAKSPVEVKSPEKAKSPVKEGAKSLAEAKSPEKAKSPVKEEI 794

Query:   355 -PPAD-ETPAEARSPLSEET-SAEEAHA-EVQSPLAEETTAEEAS--AEIQLLAAIEAPA 408
             PPA+ ++P +A+SP+ EE  S E+A +V+SP A+ EEA  A+I+ +++PA
Sbjct:   795 KPPAEVKSPEKAKSPMKEEAKSPEKAKTLDVKSPEAKTPAKEEAKRPADIRSPEQVKSPA 854

Query:   409 DETPAEAQSPLSEETSAEE-APA--EVQSPSAKGVSIEEAPLELQPPSGEETTAEEASAA 465
             E  EA+SP EET E+ AP EV+SP +EE  + +PP  E EE + A
Sbjct:   855 KE---EAKSPEKEETRTEKVAPKKEEVKSP------VEEVKAK-EPPKKVE---EEKTPA 901

Query:   466 IQLLAATEASAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAE 525
             E+ +EAP E Q P AEE + P +++P E + A+E A P E
Sbjct:   902 TPKTEVKESKKDEAPKEAQKPKAEEKEPLTEKP--KDSPGEAKKEEAKEKKAAA---PEE 956

Query:   526 EAPAEV----QPPPAEEAPAEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIEETL 581
             E PA++ + P E+A P++ PSE + P EE PA + +E E+
Sbjct:   957 ETPAKLGVKEEAKPKEKAEDAKAKEPSK--PSEKEKPKKEEVPAAPEKKDTKEEKTTESK 1014

Query:   582 AAVHSPPADDVPAEEASVDKHSPPADLL-LTEEFPIGEASAEVSPPPSEQTPEDEA 636
             P EE DK P TE+ ++ + PSE+ PED+A
Sbjct:  1015 KPEEKPKMQAKAKEE---DKGLPQEPSKPKTEKAEKSSSTDQKDSQPSEKAPEDKA 1067

 Score = 421 (63.2 bits), Expect = 3.7e-36, P = 3.7e-36
 Identities = 162/540 (30%), Positives = 275/540 (50%)

Query:   135 TEKFPAKIQPPLVEEATAKAEPR-----PAEETHVQVQPSTEETPDAEAATAVAENSVKV 189
             TE  P KI P +  K+E +  +E+ V V+  TEE  E  T  E  +
Sbjct:   419 TEGLP-KI-PSMSTHIKVKSEEKIKVVEKSEKETVIVEEQTEEIQVTEEVTE--EEDKEA 474

Query:   190 QPPPAEEAPLVEFPAEIQPPSAEESPSVELLAEILPPSAEE--SPSE-EPPAEILPPPAE 246
             Q EEA A P AEE+ S E E P EE SP+E + PAE P
Sbjct:   475 QGEEEEEAEEGGEEAATTSPPAEEAASPE--KETKSPVKEEAKSPAEAKSPAEAKSPAEA 532

Query:   247 KSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLPEEAP 306
             KSP+ E++SP+ K+P E + PAE ++PA+V+ P ++ AE + ++P
Sbjct:   533 KSPA-----EVKSPAEVKSPAEAKS-PAEA---KSPAEVKSPATVKSPAEAKSPAEAKSP 583
```

```
Query:   307 REEARELQLSTAME--TPAE-EAPTEFQSPLPKETTAEEAS-AEIQLLAATEPPAD-ETP 361
              E       + +  E  +PAE ++P E +SP+  ++ AE  S A ++      + PA+ ++P
Sbjct:   584 AEVKSPATVKSPGEAKSPAEAKSPAEVKSPVEAKSPAEAKSPASVKSPGEAKSPAEAKSP 643

Query:   362 AEARSPLSEETSAE-EAHAEVQSPLAEETTAEEASAEIQLLAAIEAPAD-ETPAEAQSPL 419
             AE +SP + ++  E ++ AEV+SP+  ++ AE A + ++    +++PA  ++P+EA+SP
Sbjct:   644 AEVKSPATVKSPVEAKSPAEVKSPVTVKSPAE-AKSPVE----VKSPASVKSPSEAKSP- 697

Query:   420 SEETSAEEAPAEVQSPS-AKGVSIEEAPLELQPPSGEETTAEEASAAIQLLAATEASAEE 478
                  +  ++PAE +SP  AK   +  ++P E +PP+  ++ AE  S A    A + A A+
Sbjct:   698 ----AGAKSPAEAKSPVVAKSPAEAKSPAEAKPPAEAKSPAEAKSPAE---AKSPAEAK- 749

Query:   479 APAEVQPPPAEEAPAEVQPPPAEEAP--AEVQPPPAEEAPA--EVQPPPAEEAPAEVQPP 534
             +PAE + P    ++P + + P  E A   AE + P   ++P   E++PP   ++P + + P
Sbjct:   750 SPAEAKSPVEVKSPEKAKSPVKEGAKSLAEAKSPEKAKSPVKEEIKPPAEVKSPEKAKSP 809

Query:   535 PAEEAPAEVQPPPAEEAPSEVQPPPAEEA--PAEVQSLPAEETPIEETLAAVHSPPADDV 592
                EEA +  +   +    E + P  EEA  PA+++S  ++P +E     SP  ++
Sbjct:   810 MKEEAKSPEKAKTLDVKSPEAKTPAKEEAKRPADIRSPEQVKSPAKEE---AKSPEKEET 866

Query:   593 PAEEASVDKHS--PPADLLLTEEFPIGEASAEVSPPPSEQTPEDEALVENVSTEFQSPQV 650
                E+ +  K       P + + +E P  +   E   P + +T   E+ +   E Q P+
Sbjct:   867 RTEKVAPKKEEVKSPVEEVKAKEPP--KKVEEEKTPATPKTEVKESKKDEAPKEAQKPKA 924

Query:   651 AGIPAVKLGSVVLEGEAKFEEVSK 674
                  +       GEAK EE  +
Sbjct:   925 EEKEPLTEKPKDSPGEAKKEEAKE 948
```

Score = 406 (60.9 bits), Expect = 1.7e-34, P = 1.7e-34
Identities = 123/390 (31%), Positives = 213/390 (54%)

```
Query:   308 EEARELQLSTAMETPAEEAPTEFQSPLPKETTAEEASAEIQLLAATEPPADETPA---EA 364
             E+  E+Q++    E   EE   E Q     +E  AEE   E    A T PPA+E   +   E
Sbjct:   455 EQTEEIQVT---EEVTEEEDKEAQGE--EEEEAEEGGEEA---ATTSPPAEEAASPEKET 506

Query:   365 RSPLSEETSAEEAHAEVQSPLAEETTAEEAS-AEIQLLAAIEAPAD-ETPAEAQSPLSEE 422
             +SP+ EE   +     AE +SP   ++ AE  S AE++  A +++PA+ ++PAEA+SP  +
Sbjct:   507 KSPVKEEAKSP---AEAKSPAEAKSPAEAKSPAEVKSPAEVKSPAEAKSPAEAKSPAEVK 563

Query:   423 TSAE-EAPAEVQSPS-AKGVSIEEAPLELQPPSGEETTAEEASAAIQLLAATEASAEEAP 480
             + A  ++PAE +SP+ AK   +  ++P  ++ P GE  + EA  +  ++ + EA   ++P
Sbjct:   564 SPATVKSPAEAKSPAEAKSPAEVKSPATVKSP-GEAKSPAEAKSPAEVKSPVEA---KSP 619

Query:   481 AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAP 540
             AE + P + ++P E + P   ++PAEV+ P   ++P E + P   ++P  V+ P   ++P
Sbjct:   620 AEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKSPVTVKSPAEAKSP 679

Query:   541 AEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIEETLAAVHSPPAD-DVPAEEASV 599
             EV+ P + ++PSE + P   ++PAE +S   ++P E   A   PPA+  PAE  S
Sbjct:   680 VEVKSPASVKSPSEAKSPAGAKSPAEAKSPVVAKSPAEAKSPAEAKPPAEAKSPAEAKSP 739

Query:   600 DKHSPPADLLLTEEFPIGEASAEVSPPPSEQTPEDEALVENVSTEFQSPQVAGIPAVKLG 659
              +   PA+    E   ++   EV   P  ++P  E  ++++ E +SP+ A  P VK
Sbjct:   740 AEAKSPAEAKSPAE---AKSPVEVKSPEKAKSPVKEG-AKSLA-EAKSPEKAKSP-VK-E 792

Query:   660 SVVLEGEAKFEEVSKINSVLKDLSNTNDGQAPTLEIES 697
              +    E K  E +K  S +K+ + + + +A TL+++S
Sbjct:   793 EIKPPAEVKSPEKAK--SPMKEEAKSPE-KAKTLDVKS 827
```

Score = 255 (38.3 bits), Expect = 5.5e-18, P = 5.5e-18
Identities = 124/420 (29%), Positives = 199/420 (47%)

```
Query:   252 ELLGEIRSPSAQKAPIEVQPLPA-----EGALEEAPAKVEPPTVEETLAEVQPLLPEEAP 306
             ELLG+I+   A +A  +  +  A        AL  E  A++E  TV+ TL    +
Sbjct:   236 ELLGQIQGCGAAQAQAQAEARDALKCDVTSALREIRAQLEGHTVQSTLQSEEWFRVRLDR 295

Query:   307 REEARELQLSTAMETPAEEAPTEFQSPLPKETTAEEASAEIQLLAATEPPADETPAEARS 366
                EA ++  + AM +  EE  TE++  L   TT      E++ L +T+   +  +E
Sbjct:   296 LSEAAKVN-TDAMRSAQEEI-TEYRRQLQARTT------ELEALKSTKESLERQRSELED 347

Query:   367 PLSEE-TSAEEAHAEVQSPLAEETTAEEASA--EIQLLAAIEAPAD-ETPAEAQSPLSEE 422
                +  S ++A  ++ + L   T  E A+  E Q L  ++  D  E  A  +  EE
Sbjct:   348 RHQVDMASYQDAIQQLDNEL-RNTKWEMAAQLREYQDLLNVKMALDIEIAAYRKLLEGEE 406

Query:   423 TSAEEAPAEV-------QSPS-AKGVSIE-EAPLELQPPSGEETT-AEEASAAIQLLA-A 471
                P+              + PS  +  + ++ E  +++     S +ET   EE  +  IQ+
Sbjct:   407 CRIGFGPSPFSLTEGLPKIPSMSTHIKVKSEEKIKVVEKSEKETVIVEEQTEEIQVTEEV 466

Query:   472 TEASAEEAPAEVQPPPAEEAPAEVQP--PPAEEAPA---EVQPPPAEEA--PAEVQPPPA 524
             TE    +EA  E +   AEE   E      PPAEEA +   E + P  EEA  PAE + P
Sbjct:   467 TEEDKEAQGE-EEEEAEEGGEEAATTSPPAEEAASPEKETKSPVKEEAKSPAEAKSPAE 525
```

```
Query:   525 EEAPAEVQPPPAEEAPAEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAE-ETPIE-ETLA 582
               ++PAE + P   ++PAEV+ P   ++P+E + P   ++PA V+S PAE ++P E ++ A
Sbjct:   526 AKSPAEAKSPAEVKSPAEVKSPAEAKSPAEAKSPAEVKSPATVKS-PAEAKSPAEAKSPA 584

Query:   583 AVHSPPADDVPAEEASVDKHSPPADLLLTEEFPIGEASAEVSPPPSEQTP-EDEALVENV 641
               V SP     P E  S  +    PA++     E     ++ AE    P S ++P E ++  E
Sbjct:   585 EVKSPATVKSPGEAKSPAEAKSPAEVKSPVE---AKSPAEAKSPASVKSPGEAKSPAEAK 641

Query:   642 S-TEFQSPQVAGIP 654
             S  E +SP     P
Sbjct:   642 SPAEVKSPATVKSP 655

 Score = 253 (38.0 bits), Expect = 9.0e-18, P = 9.0e-18
 Identities = 115/364 (31%), Positives = 166/364 (45%)

Query:   110 EVTVPVVQEGSAVKKVASAEIEPPSTEKFPAKIQPPLVEEATAKAEPRPAE-ETHVQVQ- 167
             E    PVV +   A  K + AE +PP+  K PA+ + P    ++ A+A+   PAE ++ V+V+
Sbjct:   705 EAKSPVVAKSPAEAK-SPAEAKPPAEAKSPAEAKSPAEAKSPAEAKS-PAEAKSPVEVKS 762

Query:   168 PSTEETPDAEAATAVAE--NSVKVQPPPAEEA--PL-VEFPAEIQPPSAEE--SPSVELL 220
             P    ++P  E A ++AE   +  K + P   EE   P  V+ P + + P  EE   SP
Sbjct:   763 PEKAKSPVKEGAKSLAEAKSPEKAKSPVKEEIKPPAEVKSPEKAKSPMKEEAKSPEKAKT 822

Query:   221 AEILPPSAEESPSEEP--PAEILPPPAEKSPSVELLGEIRSPSAQKAPIE-VQPLPAE-- 275
             ++  P A+     EE   PA+I  P   KSP+ E   E +SP  ++    E V  P   E
Sbjct:   823 LDVKSPEAKTPAKEEAKRPADIRSPEQVKSPAKE---EAKSPEKEETRTEKVAPKKEEVK 879

Query:   276 GALEEAPAKVEPPTVEETLAEVQPLLPEEAPREEARELQLSTAMETPAEEA-P-TEFQSP 333
              +EE   AK  P  VEE    E  P  P+    +E   ++      A +  AEE  P TE
Sbjct:   880 SPVEEVKAKEPPKKVEE---EKTPATPKTEVKESKKDEAPKEAQKPKAEEKEPLTEKPKD 936

Query:   334 LPKETTAEEASAEIQLLAATEPPADETPAE--ARSPLSEETSAEEAHA-EVQSPLAEETT 390
             P E    EEA  +    AA    P +ETPA+  +      + AE+A A E    P +E
Sbjct:   937 SPGEAKKEEAKEK---KAAA--PEEETPAKLGVKEEAKPKEKAEDAKAKEPSKPSEKEKP 991

Query:   391 A-EEASAEIQLLAAIEAPADETPAEAQSPLSEETSAEEAPAEVQSPSA-KGVSIEEAPLE 448
               EE  A  +        E    E+   + P  +  + EE     Q PS  K    E++
Sbjct:   992 KKEEVPAAPEKKDTKEEKTTESKKPEEKPKMQAKAKEEDKGLPQEPSKPKTEKAEKSSST 1051

Query:   449 LQPPSGEETTAEEASAA 465
             Q  S     A E  AA
Sbjct:  1052 DQKDSQPSEKAPEDKAA 1068
```

Pedant information for DKFZphtes3_17f10, frame 3

Report for DKFZphtes3_17f10.3

```
[LENGTH]      710
[MW]          75131.94
[pI]          4.02
[KW]          All_Alpha
[KW]          LOW_COMPLEXITY    34.08 %

SEQ   MDRSQQTSRTGYWTMMNIPPVEKVDKEQQTYFSESEIVVISRPDSSSTKSKEDALKHKSS
SEG   ............................................................
PRD   cccccccccccccccccccceeehhhhhhhccccceeeeecccccccccchhhhhhhhccc

SEQ   GKIFASEHPEFQPATNSNEEIGQKNISRTSFTQETKKGPPVLLEDELREEVTVPVVQEGS
SEG   ............................................................
PRD   cceeeccccccccccccccccccccccccccceeeecccccchhhhhhhhhheeeecccc

SEQ   AVKKVASAEIEPPSTEKFPAKIQPPLVEEATAKAEPRPAEETHVQVQPSTEETPDAEAAT
SEG   .................................................xxxxxxxxxxx
PRD   chhhhhhhccccccccccccccccchhhhhhhhhcccccceeeeccccccccccchhhhh

SEQ   AVAENSVKVQPPPAEEAPLVEFPAEIQPPSAEESPSVELLAEILPPSAEESPSEEPPAEI
SEG   xxxx...................................xxxxxxxxxxxxxxxxxxx
PRD   hhhhhccccccccccceeeecccccccccccccccchhhhhhcccccccccccccccccc

SEQ   LPPPAEKSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEPPTVEETLAEVQPL
SEG   xxxxxx.......................xxxxxxxxxxxxx..............xxx
PRD   cccccccccccccccccccccccccccccccccccchhhhhcccccccccchhhhhhhhhhc

SEQ   LPEEAPREEARELQLSTAMETPAEEAPTEFQSPLPKETTAEEASAEIQLLAATEPPADET
SEG   xxxxxxxxxxxxxx....xxxxxxxxxx....................xxxxxxxxxx
PRD   cccchhhhhhhhhhhhhhcccccccccccccccchhhhhhhhhhhhhhhhcccccccc
```

```
SEQ    PAEARSPLSEETSAEEAHAEVQSPLAEETTAEEASAEIQLLAAIEAPADETPAEAQSPLS
SEG    xxxx....xxxxxxxxxxxx.....xxxxxxxxxxxx...................xxxx
PRD    cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccc

SEQ    EETSAEEAPAEVQSPSAKGVSIEEAPLELQPPSGEETTAEEASAAIQLLAATEASAEEAP
SEG    xxxxxxxxxxx.......................xxxxxxxxxxx.......xxxxxxxx
PRD    chhhhhcccccccccccceeeccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhc

SEQ    AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAP
SEG    xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD    cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ    AEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIEETLAAVHSPPADDVPAEEASVD
SEG    xxxxxxxxxxxxxxxxxxxxxxxxxxx.................................
PRD    ccccccccccccccccccccccccccccccccccchhhhhhhhcccccccccccccccc

SEQ    KHSPPADLLLTEEFPIGEASAEVSPPPSEQTPEDEALVENVSTEFQSPQVAGIPAVKLGS
SEG    ............................................................
PRD    cccccceeeeeccccccccccccccccccccccchhhhhcccccccccccccccccccc

SEQ    VVLEGEAKFEEVSKINSVLKDLSNTNDGQAPTLEIESVFHIELKQRPPEL
SEG    ..................................................
PRD    eeeehhhhhhhhcccceeeeeeccccccccceeeehhhhhhhhhhcccccc
```

(No Prosite data available for DKFZphtes3_17f10.3)

(No Pfam data available for DKFZphtes3_17f10.3)

DKFZphtes3_17117
----

group: metabolism

DKFZphtes3_17117 encodes a novel 626 amino acid protein with similarity to transketaloases (EC 2.2.1.1).

The novel protein contains a ATP/GTP-binding site motif A (P-loop). It is a new testis-specific transketolase. Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH.

The new protein can find application in modulation of metabolic pathways involving this transketolase activity and as a new enzyme for biotechnologic production processes.

strong similarity to transketolases

few EST hits (all from testis or pooled librarys containing testis)
testis specific transketolase?

Sequenced by GBF

Locus: unknown

Insert length: 2688 bp
Poly A stretch at pos. 2649, polyadenylation signal at pos. 2630

```
   1 GACAAAAGAG AGATGATGGC CAACGACGCC AAGCCCGACG TGAAGACCGT
  51 GCAGGTGCTG CGGGACACAG CCAACCGCCT GCGGATCCAT TCCATCAGGG
 101 CCACGTGTGC CTCTGGTTCT GGCCAGCTCA CGTCGTGCTG CAGTGCAGCG
 151 GAGGTCGTGT CTGTCCTCTT CTTCCACACG ATGAAGTATA AACAGACAGA
 201 CCCAGAACAC CCGGACAACG ACCGGTTCAT CCTCTCCAGG GGACATGCTG
 251 CTCCTATCCT CTATGCTGCT TGGGTGGAGG TGGGTGACAT CAGTGAATCT
 301 GACTTGCTGA ACCTGAGGAA ACTTCACAGC GACTTGGAGA GACACCCTAC
 351 CCCGCGATTG CCGTTTGTTG ACGTGGCAAC AGGGTCCCTA GGTCAGGGAT
 401 TAGGTACTGC ATGTGGAATG GCTTATACTG GCAAGTACCT TGACAAGGCC
 451 AGCTACCGGG TGTTCTGCCT TATGGGAGAT GGCGAATCCT CAGAAGGCTC
 501 TGTGTGGGAG GCTTTTGCTT TTGCCTCCCA CTACAACTTG GACAATCTCG
 551 TGGCGGTCTT CGACGTGAAC CGCTTGGGAC AAAGTGGCCC TGCACCCCTT
 601 GAGCATGGCG CAGACATCTA CCAGAATTGC TGTGAAGCCT TTGGATGGAA
 651 TACTTACTTA GTGGATGGCC ATGATGTGGA GGCCTTGTGC CAAGCATTTT
 701 GGCAAGCAAG TCAAGTGAAG AACAAGCCTA CTGCTATAGT TGCCAAGACC
 751 TTCAAAGGTC GGGGTATTCC AAATATTGAG GATGCAGAAA ATTGGCATGG
 801 AAAGCCAGTG CCAAAAGAAA GAGCAGATGC AATTGTCAAA TTAATTGAGA
 851 GTCAGATACA GACCAATGAG AATCTCATAC CAAAATCGCC TGTGGAAGAC
 901 TCACCTCAAA TAAGCATCAC AGATATAAAA ATGACCTCCC CACCTGCTTA
 951 CAAAGTTGGT GACAAGATAG CTACTCAGAA AACATATGGT TTGGCTCTGG
1001 CTAAACTGGG CCGTGCAAAT GAAAGAGTTA TTGTTCTGAG TGGTGACACG
1051 ATGAACTCCA CCTTTTCTGA GATATTCAGG AAAGAACACC CTGAGCGTTT
1101 CATAGAGTGT ATTATTGCTG AACAAAACAT GGTAAGTGTG GCACTAGGCT
1151 GTGCTACACG TGGTCGAACC ATTGCTTTTG CTGGTGCTTT TGCTGCCTTT
1201 TTTACTAGAG CATTCGATCA GCTCCGAATG GGAGCCATTT CTCAAGCCAA
1251 TATCAACCTT ATTGGTTCCC ACTGTGGGGT ATCCACTGGA GAAGATGGAG
1301 TCTCCCAGAT GGCCCTGGAG GATCTAGCCA TGTTCCGAAG CATTCCCAAT
1351 TGTACTGTTT TCTATCCAAG TGATGCCATC TCGACAGAGC ATGCTATTTA
1401 TCTAGCCGCC AATACCAAGG GAATGTGCTT CATTCGAACC AGCCAACCAG
1451 AAACTGCAGT TATTTATACC CCACAAGAAA ATTTTGAGAT TGGCCAGGCC
1501 AAGGTGGTCC GCCACGGTGT CAATGATAAA GTCACAGTAA TTGGAGCTGG
1551 AGTTACTCTC CATGAAGCCT TAGAAGCTGC TGACCATCTT TCTCAACAAG
1601 GTATTTCTGT CCGTGTCATC GACCCATTTA CCATTAAACC CCTGGATGCC
1651 GCCACCATCA TCTCCAGTGC AAAAGCCACA GGCGGCCGAG TTATCACAGT
1701 GGAGGATCAC TACAGGGAAG GTGGCATTGG AGAAGCTGTT TGTGCAGCTG
1751 TCTCCAGGGA GCCTGATATC CTTGTTCATC AACTGGCAGT GTCAGGAGTG
1801 CCTCAACGTG GGAAAACTAG TGAATTGCTG GATATGTTTG GAATCAGTAC
1851 CAGACACATT ATAGCAGCCG TAACACTTAC TTTAATGAAG TAAACTAGGC
1901 TTATTTCTAA AAAGTCAAGT CTATTGGCTT TGGCCCAAAA GCACTGGTAT
1951 CTTTGTATTA AATTCATGTT TATTGTCACA AAACCATTAT TTATACCTAT
2001 ACAGTTGTAC TGTTTCTTTT AAAGCAAAGC CATTTAACAT CTTTCTTCAT
2051 TCCTAATTTG GAAATTAAAG TTTACCTTTC TGTTAATCTA TGTATAAATG
2101 TTACTCTGAG TTATTAATGT GGATTTTAAA ATTGTAAGCA ATAGAATAGG
2151 AAATAAAACA ACTACCTAAT ACAAATATTT CTGATAAGAC TACAAATATC
2201 TGACTGAGCT GGGGATTAAA GTAGAGGTAA CTGTATCTTA AATGAGTATG
2251 ATTTCCTTGT AAGTTAAAAA AATTGAAATT TAATTGTAGA CTTCAATAGT
2301 CCAAGTTTTG AAGGATGTTT GAGCTTTTGT ATAATGCCAT TTATACCTGC
2351 AGTTTTACAG ATAATGTTTG ACTGCAGTTG CCTTGGAAAT TCCTCCAAAG
2401 TTTGCCTTCA TCTCTCCTCT ACAGTTTGGA GGTGATGGTG CAGCAGTGGA
2451 ACATCTCTTG ATGCACCACA CTACTTGTGT TCTGTGAAGT GATGAAAGTA
```

```
2501 TAACTGGTTC TAGTTTGCAC ACTACACACA TAGTTTTGTG AAGCTTCAGA
2551 AATGTTTTTT CTTTTCCTTG TGGCCAAACC AGTTTGTTAA TCTGATTATA
2601 TTCATCTGCT AATGATACTA AAGTTAATGT AATAAAGCAT TTAAAAATCA
2651 GAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

96214928:
Amplification of the transketolase gene in desensitization-resistant mutant
Y1 mouse adrenocortical tumor cells.

99123875:
Properties and functions of the thiamin diphosphate dependent enzyme transketolase.

Peptide information for frame 1

ORF from 13 bp to 1890 bp; peptide length: 626
Category: strong similarity to known protein
Classification: Metabolism
Prosite motifs: ATP_GTP_A (595-603)

```
  1 MMANDAKPDV KTVQVLRDTA NRLRIHSIRA TCASGSGQLT SCCSAAEVVS
 51 VLFFHTMKYK QTDPEHPDND RFILSRGHAA PILYAAWVEV GDISESDLLN
101 LRKLHSDLER HPTPRLPFVD VATGSLGQGL GTACGMAYTG KYLDKASYRV
151 FCLMGDGESS EGSVWEAFAF ASHYNLDNLV AVFDVNRLGQ SGPAPLEHGA
201 DIYQNCCEAF GWNTYLVDGH DVEALCQAFW QASQVKNKPT AIVAKTFKGR
251 GIPNIEDAEN WHGKPVPKER ADAIVKLIES QIQTNENLIP KSPVEDSPQI
301 SITDIKMTSP PAYKVGDKIA TQKTYGLALA KLGRANERVI VLSGDTMNST
351 FSEIFRKEHP ERFIECIIAE QNMVSVALGC ATRGRTIAFA GAFAAFFTRA
401 FDQLRMGAIS QANINLIGSH CGVSTGEDGV SQMALEDLAM FRSIPNCTVF
451 YPSDAISTEH AIYLAANTKG MCFIRTSQPE TAVIYTPQEN FEIGQAKVVR
501 HGVNDKVTVI GAGVTLHEAL EAADHLSQQG ISVRVIDPFT IKPLDAATII
551 SSAKATGGRV ITVEDHYREG GIGEAVCAAV SREPDILVHQ LAVSGVPQRG
601 KTSELLDMFG ISTRHIIAAV TLTLMK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17117, frame 1

SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68)., N = 1, Score = 2222, P = 2.5e-230

SWISSPROT:TKT_RAT TRANSKETOLASE (EC 2.2.1.1) (TK)., N = 1, Score = 2202, P = 3.3e-228

TREMBL:RN09256_1 product: "transketolase"; Rattus norvegicus Sprague-Dawley transketolase mRNA, complete cds., N = 1, Score = 2202, P = 3.3e-228

SWISSPROT:TKT_HUMAN TRANSKETOLASE (EC 2.2.1.1) (TK)., N = 1, Score = 2200, P = 5.3e-228

>SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).
Length = 623

HSPs:

Score = 2222 (333.4 bits), Expect = 2.5e-230, P = 2.5e-230
Identities = 417/614 (67%), Positives = 501/614 (81%)

```
Query:     7 KPDVKTVQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLFFHTMKYKQTDPEH 66
```

```
            KPD + +Q L+DTANRLRI SI+AT A+GSG  TSCCSAAE+++VLFFHTM+YK  DP +
Sbjct:    6 KPDQQKLQALKDTANRLRISSIQATTAAGSGHPTSCCSAAEIMAVLFFHTMRYKALDPRN 65

Query:   67 PDNDRFILSRGHAAPILYAAWVEVGDISESDLLNLRKLHSDLERHPTPRLPFVDVATGSL 126
            P NDRF+LS+GHAAPILYA W E G + E++LLNLRK+ SDL+ HP P+  F DVATGSL
Sbjct:   66 PHNDRFVLSKGHAAPILYAVWAEAGFLPEAELLNLRKISSDLDGHPVPKQAFTDVATGSL 125

Query:  127 GQGLGTACGMAYTGKYLDKASYRVFCLMGDGESSEGSVWEAFAFASHYNLDNLVAVFDVN 186
            GQGLG ACGMAYTGKY DKASYRV+C++GDGE SEGSVWEA AFA  Y LDNLVA+FD+N
Sbjct:  126 GQGLGAACGMAYTGKYFDKASYRVYCMLGDGEVSEGSVWEAMAFAGIYKLDNLVAIFDIN 185

Query:  187 RLGQSGPAPLEHGADIYQNCCEAFGWNTYLVDGHDVEALCQAFWQASQVKNKPTAIVAKT 246
            RLGQS PAPL+H  DIYQ  CEAFGW+T +VDGH VE LC+AF QA    K++PTAI+AKT
Sbjct:  186 RLGQSDPAPLQHQVDIYQKRCEAFGWHTIIVDGHSVEELCKAFGQA---KHQPTAIIAKT 242

Query:  247 FKGRGIPNIEDAENWHGKPVPKERADAIVKLIESQIQTNENLIPKSPVEDSPQISITDIK 306
            FKGRGI  IED E WHGKP+PK  A+ I++ I SQ+Q+ + ++   P ED+P + I +I+
Sbjct:  243 FKGRGITGIEDKEAWHGKPLPKNMAEQIIQEIYSQVQSKKKILATPPQEDAPSVDIANIR 302

Query:  307 MTSPPAYKVGDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKEHPERFIEC 366
            M +PP+YKVGDKIAT+K YGLALAKLG A++R+I L GDT NSTFSE+F+KEHP+RFIEC
Sbjct:  303 MPTPPSYKVGDKIATRKAYGLALAKLGHASDRIIALDGDTKNSTFSELFKKEHPDRFIEC 362

Query:  367 IIAEQNMVSVALGCATRGRTIAFAGAFAAFFTRAFDQLRMGAISQANINLIGSHCGVSTG 426
             IAEQNMVS+A+GCATR RT+ F   FAAFFTRAFDQ+RM AIS++NINL GSHCGVS G
Sbjct:  363 YIAEQNMVSIAVGCATRDRTVPFCSTFAAFFTRAFDQIRMAAISESNINLCGSHCGVSIG 422

Query:  427 EDGVSQMALEDLAMFRSIPNCTVFYPSDAISTEHAIYLAANTKGMCFIRTSQPETAVIYT 486
            EDG SQMALEDLAMFRS+P  TVFYPSD ++TE A+ LAANTKG+CFIRTS+PE A+IY+
Sbjct:  423 EDGPSQMALEDLAMFRSVPMSTVFYPSDGVATEKAVELAANTKGICFIRTSRPENAIIYS 482

Query:  487 PQENFEIGQAKVVRHGVNDKVTVIGAGVTLHEALEAADHLSQQGISVRVIDPFTIKPLDA 546
             E+F++GQAKVV     +D+VTVIGAGVTLHEAL AA+  L +  IS+RV+DPFTIKPLD
Sbjct:  483 NNEDFQVGQAKVVLKSKDDQVTVIGAGVTLHEALAAAESLKKDKISIRVLDPFTIKPLDR 542

Query:  547 ATIISSAKATGGRVITVEDHYREGGIGEAVCAAVSREPDILVHQLAVSGVPQRGKTSELL 606
             I+ SA+AT GR++TVEDHY EGGIGEAV AAV  EP + V +LAVS VP+ GK +ELL
Sbjct:  543 KLILDSARATKGRILTVEDHYYEGGIGEAVSAAVVGEPGVTVTRLAVSQVPRSGKPAELL 602

Query:  607 DMFGISTRHIIAAV 620
             MFGI    I+ AV
Sbjct:  603 KMFGIDKDAIVQAV 616
```

Pedant information for DKFZphtes3_17117, frame 1

Report for DKFZphtes3_17117.1

```
[LENGTH]     626
[MW]         67877.52
[pI]         5.90
[HOMOL]      SWISSPROT:TKT_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68). 0.0
[FUNCAT]     m outer membrane and cell wall      [M. jannaschii, MJ0681] 3e-48
[FUNCAT]     g carbohydrate metabolism and transport      [H. influenzae, HI1023] 9e-36
[FUNCAT]     01.05.01 carbohydrate utilization    [S. cerevisiae, YPR074c] 5e-32
[FUNCAT]     30.03 organization of cytoplasm      [S. cerevisiae, YPR074c] 5e-32
[FUNCAT]     02.07 pentose-phosphate pathway      [S. cerevisiae, YPR074c] 5e-32
[FUNCAT]     01.01.01 amino-acid biosynthesis     [S. cerevisiae, YPR074c] 5e-32
[FUNCAT]     i lipid metabolism     [H. influenzae, HI1439] 3e-17
[FUNCAT]     c energy conversion    [H. influenzae, HI1233] 2e-09
[FUNCAT]     02.01 glycolysis       [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase]
2e-05
[FUNCAT]     30.16 mitochondrial organization     [S. cerevisiae, YBR221c PDB1 - pyruvate
dehydrogenase] 2e-05
[BLOCKS]     BL00801F
[BLOCKS]     BL00801E
[BLOCKS]     BL00801D Transketolase proteins
[BLOCKS]     BL00801C Transketolase proteins
[BLOCKS]     BL00801B Transketolase proteins
[BLOCKS]     BL00801A Transketolase proteins
[SCOP]       d1trka2 3.28.1.2.1 Transketolase Transketolase, C-terminal domai 1e-21
[EC]         1.2.4.1 Pyruvate dehydrogenase (lipoamide) 8e-11
[EC]         1.2.4.4 3-Methyl-2-oxobutanoate dehydrogenase (lipoamide) 4e-10
[EC]         2.2.1.1 Transketolase 0.0
[EC]         2.2.1.3 Formaldehyde transketolase 1e-20
[PIRKW]      transferase 0.0
[PIRKW]      flavoprotein 2e-07
[PIRKW]      Calvin cycle 1e-40
[PIRKW]      heterotetramer 2e-07
```

```
[PIRKW]        pentose phosphate pathway 0.0
[PIRKW]        magnesium 1e-40
[PIRKW]        thiamine pyrophosphate 0.0
[PIRKW]        oxidoreductase 7e-12
[PIRKW]        fatty acid biosynthesis 4e-10
[PIRKW]        mitochondrion 2e-07
[PIRKW]        peroxisome 1e-20
[PIRKW]        homodimer 1e-40
[SUPFAM]       pyruvate dehydrogenase (lipoamide) alpha chain 1e-06
[SUPFAM]       pyruvate dehydrogenase (lipoamide) beta chain 7e-12
[SUPFAM]       ferredoxin 2[4Fe-4S]-related protein 8e-47
[SUPFAM]       thiamine pyrophosphate-binding domain homology 0.0
[SUPFAM]       pyruvate dehydrogenase (lipoamide) 6e-08
[SUPFAM]       ferredoxin 2[4Fe-4S] homology 8e-47
[SUPFAM]       hypothetical protein C2814 2e-21
[SUPFAM]       transketolase 0.0
[PROSITE]      ATP_GTP_A       1
[PFAM]         Transketolase
[KW]           Alpha_Beta
[KW]           3D
[KW]           LOW_COMPLEXITY     3.04 %

SEQ    MMANDAKPDVKTVQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLFFHTMKYK
SEG    ............................................................
1ngsB  ......................HHHHHHHHHHHCCCCHHHHHHHHHHHHHHHH-HHCCCT

SEQ    QTDPEHPDNDRFILSRGHAAPILYAAWVEVGDISESDLLNLRKLHSDLERHPTPRLPFVD
SEG    ............................................................
1ngsB  TTTTTTTTTCEEEETTGGGHHHHHHHHHHHCTTCHHHHHTTTTTTTTTTTTTTTTTTTTC

SEQ    VATGSLGQGLGTACGMAYTGKYLDKASYRVFCLMGDGESSEGSVWEAFAFASHYNLDNLV
SEG    ............................................................
1ngsB  CCCCTTTHHHHHHHHHHHHHHHHCBTTBTTEEEECHHHHHCHHHHHHHHHHHHHCTTTEE

SEQ    AVFDVNRLGQSGPAPLEHGADIYQNCCEAFGWNTYLVDGHDVEALCQAFWQASQVKNKPT
SEG    ............................................................
1ngsB  EEEEECCEETTEEGGGCCCCCHHHHH-HHHCCEEEETTTTTHHHHHHHHHHHHHHTTTTCE

SEQ    AIVAKTFKGRGIPNIEDAENWHGKPVPKERADAIVKLIESQIQTNENLIPKSPVEDSPQI
SEG    ............................................................
1ngsB  EEEEECTTTTTTCCHHHHHHHHHHTCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHH

SEQ    SITDIKMTSPPAYKVGDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKEHP
SEG    ............................................................
1ngsB  HHHHHHHHHTCCCTTTTCBCHHHHHHHHHHHHHHTTTTEEEEETTTHHHHCCTTCEECCG

SEQ    ERFIECIIAEQNMVSVALGCATRGRTIAFAGAFAAFFTRAFDQLRMGAISQANINLIGSH
SEG    ........................xxxxxxxxxxxxxxxxxx..................
1ngsB  GCEEETTTTHHHHHHHHHHHHHHTTTTEEEEEEGGGGGGHHHHHHHHHHCTTTEEEEEC

SEQ    CGVSTGEDGVSQMALEDLAMFRSIPNCTVFYPSDAISTEHAIYLAANTKGMCFIRTSQPE
SEG    ............................................................
1ngsB  CCGGGTTTTTTTTCCHHHHHHHCTTTTEEECCCCHHHHHHHHHHHTTTTCEEEECCCCCB

SEQ    TAVIYTPQENFEIGQAKVVRHGVNDKVTVIGAGVTLHEALEAADHLSQQGISVRVIDPFT
SEG    ............................................................
1ngsB  CCTTTTCHHHHHCC-CEEEETTTTTTEEEEECCHHHHHHHHHHHHHHHHHCCCEEEE....

SEQ    IKPLDAATIISSAKATGGRVITVEDHYREGGIGEAVCAAVSREPDILVHQLAVSGVPQRG
SEG    ............................................................
1ngsB  ............................................................

SEQ    KTSELLDMFGISTRHIIAAVTLTLMK
SEG    ..........................
1ngsB  ..........................
```

Prosite for DKFZphtes3_17117.1

PS00017	595->603	ATP_GTP_A	PDOC00017

Pfam for DKFZphtes3_17117.1

HMM_NAME Transketolase

HMM *vNtIRiLaMDAVEKANSGHPGaPMGMAPMAHVLWqrMMRHNPNDPrWPN

```
                 +N++RI ++ A  + +SG  ++++++A++  VL++++M+++++DP  P+
Query        20  ANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLFFHTMKYKQTDPEHPD   68

HMM              RDRFVLSNGHaCMLLYsMWHLyGYDMpMWDLkQFRQWHSrTPGHPEIgHT
                 +DRF+LS GHA+++LY+ W + G  ++++DL+++R++HS++ +HP ++
Query        69  NDRFILSRGHAAPILYAAWVEVGD-ISESDLLNLRKLHSDLERHPTPRLP  117

HMM              PGVEVTTGPLGQGIaNaVWMAIAERnLAATYNRPGFDIfDHYTYCFMGDG
                 ++ +V+TG+LGQG++            +++++Y++++ D+++++++C+MGDG
Query       118  FV-DVATGSLGQGLG---------TACGMAYTGKYLDKASYRVFCLMGDG  157

HMM              CLMEGISWEACSLAGHMqLGNWIaFYDDNrISIDGdTdIWFqEDtYakRF
                 + +EG++WEA ++A+H++L+N++A +D NR++++G++++ +  D+Y+  +
Query       158  ESSEGSVWEAFAFASHYNLDNLVAVFDVNRLGQSGPAPLEHGADIYQNCC  207

HMM              EAYGWHVIEVEnDGHDvEeIcaAIEeAKaekDRPTLIiCRTVIGYGSPNk
                 EA+GW++ +V  DGHDVE++C A+ +A  +K++PT+I ++T++G+G+PN
Query       208  EAFGWNTYLV--DGHDVEALCQAFWQASQVKNKPTAIVAKTFKGRGIPNI  255

HMM              QGTHdWHGAPLGeD*
                 ++ + WHG+P +++
Query       256  EDAENWHGKPVPKE    269

HMM              *PqWePnddkIATRKASQqaLeaiGPaLPEfWGGSADLTPSNLTrWKGmv
                  P++++ +DKIAT K+++ AL+++G A  +++ +S+D+ +S++++++ ++
Query       311   PAYKV-GDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKE  358

HMM              WFMPPSISTDCynGNWsGRYIHYGIREHgMgAIMNGIAlHGgNFRPYGGT
                 +               + R+I++ I+E++M++++ G+A++G+ ++++ G
Query       359  H---------------PERFIECIIAEQNMVSVALGCATRGR-TIAFAGA  392

HMM              FMMFyDYARPAIRMAALMelPVIWVWTHDSIGLGEDGPTHQPVEHLAHFR
                 F++F+++A++++RM A++  +++++++H++++ GEDG +++++E+LA+FR
Query       393  FAAFFTRAFDQLRMGAISQANINLIGSHCGVSTGEDGVSQMALEDLAMFR  442

HMM              aIPNMsVWRPCDgNETayAWylAvERehTPtiLILSRQNLPQlErNPrqf
                 +IPN +V++P+D+  T+ A YLA+++++  +++++S ++ +++++ P +
Query       443  SIPNCTVFYPSDAISTEHAIYLAANTKGM-CFIRTSQPETAVIYT-PQEN  490

HMM              ekvaRGGYVLkDmdnePDVILIATGSEMELAvaAAKlLadEGIkaRVVSM
                 ++++++++V + +    + V++I++G+++++A++AA+ L+ +GI +RV+++
Query       491  FEIGQAKVVRHGVN--DKVTVIGAGVTLHEALEAADHLSQQGISVRVIDP  538

HMM              PCTeWFD.......kQDeEYReSVLPDhVPqRVaVEmGvtWCWYKYVGqq
                 ++++++D         + ++++R  +++DH++ +++++++V  ++ +++ +
Query       539  FTIKPLDAATIISSAKATGGRVITVEDHYR-EGGIGEAVCAAVSREPDIL  587

HMM              GaIfGMNrFGESSGKAPpevLYkMFGFTPENI*
                 +  +++ +++       ++  +L+ MFG+   +I
Query       588  VHQLAVSGVPQR---GKTSELLDMFGISTRHI    616
```

DKFZphtes3_17n12
----

group: transcription factors

DKFZphtes3_17n12.1 encodes a novel 804 amino acid protein which is nearly identical to mouse and trout SOX-LZ.

Sox proteins belong to the HMG box superfamily of DNA-binding proteins and are involved in the regulation of developmental processes as germ layer formation, organ development and cell type specification. Deletion or mutation of Sox proteins often results in developmental defects and congenital disease in humans. Sox proteins perform their function in a complex interplay with other transcription factors in a manner highly dependent on cell type and promoter context. The new protein is related to the SOX-LZ protein and contains an additional leucin-zipper.

The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

nearly identical to mouse SOX-LZ

complete cDNA, complete cds, few EST hits
mouse and trout SOX-LZ, involved in spermatogenesis

Sequenced by GBF

Locus: unknown

Insert length: 2802 bp
Poly A stretch at pos. 2692, polyadenylation signal at pos. 2660

```
   1 GGGATAGGAA AGATGAAAGG TCATGGTGAG CTTCAAGGAC ATGAAAGGTT
  51 GTTGTCTCAT GTAACAATAG TAGATTGTTT TTTTTCCTAA TATTTCTAGC
 101 CAGCCCCTAA GTCAGGTGAT GGAACAAATA CCTACAGTTT AGTCAGGTGA
 151 AACAGGAGTG GGTGGAGGAA GGAAAGAAGA AAAATGGGAA GAATGTCTTC
 201 CAAGCAAGCC ACCTCTCCAT TTGCCTGTGC AGCTGATGGA GAGGATGCAA
 251 TGACCCAGGA TTTAACCTCA AGGGAAAAGG AAGAGGGCAG TGATCAACAT
 301 GTGGCCTCCC ATCTGCCTCT GCACCCCATA ATGCACAACA AACCTCACTC
 351 TGAGGAGCTA CCAACACTTG TCAGTACCAT TCAACAAGAT GCTGACTGGG
 401 ACAGCGTTCT GTCATCTCAG CAAAGAATGG AATCAGAGAA TAATAAGTTA
 451 TGTTCCCTAT ATTCCTTCCG AAATACCTCT ACCTCACCAC ATAAGCCTGA
 501 CGAAGGGAGT CGGGACCGTG AGATAATGAC CAGTGTTACT TTTGGAACCC
 551 CAGAGCGCCG CAAAGGGAGT CTTGCCGATG TGGTGGACAC ACTGAAACAG
 601 AAGAAGCTTG AGGAAATGAC TCGGACTGAA CAAGAGGATT CCTCCTGCAT
 651 GGAAAAACTA CTTTCAAAAG ATTGGAAGGA AAAAATGGAA AGACTAAATA
 701 CCAGTGAACT TCTTGGAGAA ATTAAAGGTA CACCTGAGAG CCTGGCAGAA
 751 AAAGAACGGC AGCTCTCCAC CATGATTACC CAGCTGATCA GTTTACGGGA
 801 GCAGCTACTG GCAGCGCATG ATGAACAGAA AAAACTGGCA GCGTCACAAA
 851 TTGAGAAACA ACGGCAGCAA ATGGACCTTG CTCGCCAACA GCAAGAACAG
 901 ATTGCGAGAC AACAGCAGCA ACTTCTGCAA CAGCAGCACA AAATTAATCT
 951 CCTGCAGCAA CAGATCCAGG TTCAGGGTCA CATGCCTCCG CTCATGATCC
1001 CAATTTTTCC ACATGACCAG CGGACTCTGG CAGCAGCTGC TGCTGCCCAA
1051 CAGGGATTCC TCTTCCCCCC TGGAATAACA TACAAACCAG GTGATAACTA
1101 CCCCGTACAG TTCATTCCAT CAACAATGGC AGCTGCTGCT GCTTCTGGAC
1151 TCAGCCCTTT ACAGCTCCAG CAGCTCTATG CCGCTCAGCT GGCCAGCATG
1201 CAGGTGTCAC CTGGAGCAAA GATGCCATCA ACTCCACAGC CACCAAACAC
1251 AGCAGGGACG GTCTCACCTA CTGGGATAAA AAATGAAAAG AGAGGGACCA
1301 GCCCTGTAAC TCAAGTTAAG GATGAAGCAG CAGCACAGCC TCTGAATCTC
1351 TCATCCCGAC CCAAGACAGC AGAGCCTGTA AAGTCCCCAA CGTCTCCCAC
1401 CCAGAACCTC TTCCCAGCCA GCAAAACCAG CCCTGTCAAT CTGCCAAACA
1451 AAAGCAGCAT CCCTAGCCCC ATTGGAGGAA GCCTGGGAAG AGGATCCTCT
1501 TTAGGTAAAT GGAAAAGTCA ACACCAGGAA GAGACTTACG AATTAGATAT
1551 CCTATCTAGT CTCAACTCCC CTGCCCTTTT TGGGGATCAG GATACAGTGA
1601 TGAAAGCCAT TCAGGAGGCG CGGAAGATGC GAGAGCAGAT CCAGCGGGAG
1651 CAACAGCAGC AACAGCCACA TGGTGTTGAC GGGAAACTGT CCTCCATAAA
1701 TAATATGGGG CTGAACAGCT GCAGGAATGA AAAGGAAAGA ACGCGCTTTG
1751 AGAATTTGGG GCCCCAGTTA ACGGGAAAGT CAAATGAAGA TGGAAAACTG
1801 GGCCCAGGTG TCATCGACCT TACTCGGCCA GAAGATGCAG AGGGAAGTAA
1851 AGCAATGAAT GGCTCTGCAG CTAAACTACA GCAGTATTAT TGTTGGCCAA
1901 CAGGAGGTGC CACTGTGGCT GAAGCACGAG TCTACAGGGA CGCCCCGCGGC
1951 CGTGCCAGCA GCGAGCCACA CATTAAGCGA CCAATGAATG CATTCATGGT
2001 TTGGGCAAAG GATGAGAGGA GAAAAATCCT TCAGGCCTTC CCCGACATGC
2051 ATAACTCCAA CATTAGCAAA ATCTTAGGAT CTCGCTGGAA ATCAATGTCC
2101 AACCAGGAGA AGCAACCTTA TTATGAAGAG CAGGCCCGGC TAAGCAAGAT
2151 CCACTTAGAG AAGTACCCAA ACTATAAATA CAAACCCCGA CCGAAACGCA
2201 CCTGCATTGT TGATGGCAAA AAGCTTCGGA TTGGGGAGTA TAAGCAACTG
2251 ATGAGGTCTC GGAGACAGGA GATGAGGCAG TTCTTTACTG TGGGGCAACA
2301 GCCTCAGATT CCAATCACCA CAGGAACAGG TGTTGTGTAT CCTGGTGCTA
2351 TCACTATGGC AACTACCACA CCATCGCCTC AGATGACATC TGACTGCTCT
```

```
2401 AGCACCTCGG CCAGCCCGGA GCCCAGCCTC CCGGTCATCC AGAGCACTTA
2451 TGGTATGAAG ACAGATGGCG GAAGCCTAGC TGGAAATGAA ATGATCAATG
2501 GAGAGGATGA AATGGAAATG TATGATGACT ATGAAGATGA CCCCAAATCA
2551 GACTATAGCA GTGAAAATGA AGCCCCGGAG GCTGTCAGTG CCAACTGAGG
2601 AGTTTTTGTT TGCTGAATTA AAGTACTCTG ACATTTCACC CCCCTCCCCA
2651 ACAAAGAGTT ATTAAAGAGC CCGCATGCAT TTGTGGCTCC ACAATTAAAA
2701 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AA
```

BLAST Results

No BLAST result

Medline entries

95311974:
A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein.

96032826:
The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse.

Peptide information for frame 1

ORF from 184 bp to 2595 bp; peptide length: 804
Category: strong similarity to known protein

```
  1 MGRMSSKQAT SPFACAADGE DAMTQDLTSR EKEEGSDQHV ASHLPLHPIM
 51 HNKPHSEELP TLVSTIQQDA DWDSVLSSQQ RMESENNKLC SLYSFRNTST
101 SPHKPDEGSR DREIMTSVTF GTPERRKGSL ADVVDTLKQK KLEEMTRTEQ
151 EDSSCMEKLL SKDWKEKMER LNTSELLGEI KGTPESLAEK ERQLSTMITQ
201 LISLREQLLA AHDEQKKLAA SQIEKQRQQM DLARQQQEQI ARQQQQLLQQ
251 QHKINLLQQQ IQVQGHMPPL MIPIFPHDQR TLAAAAAAQQ GFLFPPGITY
301 KPGDNYPVQF IPSTMAAAAA SGLSPLQLQQ LYAAQLASMQ VSPGAKMPST
351 PQPPNTAGTV SPTGIKNEKR GTSPVTQVKD EAAAQPLNLS SRPKTAEPVK
401 SPTSPTQNLF PASKTSPVNL PNKSSIPSPI GGSLGRGSSL GKWKSQHQEE
451 TYELDILSSL NSPALFGDQD TVMKAIQEAR KMREQIQREQ QQQQPHGVDG
501 KLSSINNMGL NSCRNEKERT RFENLGPQLT GKSNEDGKLG PGVIDLTRPE
551 DAEGSKAMNG SAAKLQQYYC WPTGGATVAE ARVYRDARGR ASSEPHIKRP
601 MNAFMVWAKD ERRKILQAFP DMHNSNISKI LGSRWKSMSN QEKQPYYEEQ
651 ARLSKIHLEK YPNYKYKPRP KRTCIVDGKK LRIGEYKQLM RSRRQEMRQF
701 FTVGQQPQIP ITTGTGVVYP GAITMATTTP SPQMTSDCSS TSASPEPSLP
751 VIQSTYGMKT DGGSLAGNEM INGEDEMEMY DDYEDDPKSD YSSENEAPEA
801 VSAN
```

BLASTP hits

Entry MMSOXLZ2_1 from database TREMBL:
product: "SOX-LZ"; Mouse mRNA for SOX-LZ, complete cds.
Score = 3910, P = 0.0e+00, identities = 764/801, positives = 774/801

Entry I51083 from database PIR:
SOX-LZ - rainbow trout
Score = 1774, P = 1.1e-287, identities = 365/532, positives = 431/532

Entry S59121 from database PIR:
SOX6 protein - mouse
Score = 2319, P = 1.2e-240, identities = 489/660, positives = 527/660

Entry AB006330_1 from database TREMBL:
gene: "mSox5L"; product: "SOX5"; Mus musculus mSox5L mRNA, complete cds.
Score = 1212, P = 8.9e-209, identities = 274/457, positives = 324/457

Entry MMU010604_1 from database TREMBL:
gene: "sox5"; product: "L-Sox5 protein"; Mus musculus mRNA for transcription factor L-Sox5
Score = 879, P = 4.2e-195, identities = 190/281, positives = 218/281

Alert BLASTP hits for DKFZphtes3_17n12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_17n12, frame 1

Report for DKFZphtes3_17n12.1

```
[LENGTH]    804
[MW]        89332.69
[pI]        6.97
[HOMOL]     TREMBL:MMSOXLZ2_1 product: "SOX-LZ"; Mouse mRNA for SOX-LZ, complete cds. 0.0
[FUNCAT]    04.05.01.04 transcriptional control  [S. cerevisiae, YKL032c] 8e-07
[FUNCAT]    30.10 nuclear organization   [S. cerevisiae, YKL032c] 8e-07
[FUNCAT]    01.07.07 regulation of vitamins, cofactors, and prosthetic groups  [S.
cerevisiae, YPR065w] 5e-06
[FUNCAT]    03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR089c-a]
7e-06
[FUNCAT]    30.13 organization of chromosome structure  [S. cerevisiae, YBR089c-a] 7e-06
[FUNCAT]    03.01 cell growth    [S. cerevisiae, YBR089c-a] 7e-06
[FUNCAT]    03.16 dna synthesis and replication  [S. cerevisiae, YMR072w] 2e-04
[FUNCAT]    30.16 mitochondrial organization    [S. cerevisiae, YMR072w] 2e-04
[SCOP]      d1hmf__ 1.20.1.1.1 HMG1, fragments A and B [rat/hamster (Rattu 1e-13
[SCOP]      d1lefa_ 1.20.1.1.6 Lymphoid enhancer-binding factor, LEF1 [mous 4e-15
[SCOP]      d1hrya_ 1.20.1.1.4 SRY [Human (Homo sapiens) 7e-17
[PIRKW]     DNA binding 4e-94
[PIRKW]     T-cell receptor 4e-07
[PIRKW]     leucine zipper 1e-38
[PIRKW]     alternative splicing 2e-07
[PIRKW]     transcription factor 4e-16
[PIRKW]     transcription regulation 1e-12
[SUPFAM]    HMG box homology 0.0
[SUPFAM]    unassigned HMG box proteins 4e-94
[PROSITE]   ATP_GTP_A       1
[PROSITE]   LEUCINE_ZIPPER 1
[PROSITE]   MYRISTYL        6
[PROSITE]   AMIDATION       1
[PROSITE]   CAMP_PHOSPHO_SITE    2
[PROSITE]   CK2_PHOSPHO_SITE     14
[PROSITE]   PKC_PHOSPHO_SITE     10
[PROSITE]   ASN_GLYCOSYLATION    6
[PFAM]      HMG (high mobility group) box
[KW]        Irregular
[KW]        3D
[KW]        LOW_COMPLEXITY    13.81 %
[KW]        COILED_COIL        3.48 %
```

```
SEQ     MGRMSSKQATSPFACAADGEDAMTQDLTSREKEEGSDQHVASHLPLHPIMHNKPHSEELP
SEG     ............................................................
COILS   ............................................................
1nhm-   ............................................................

SEQ     TLVSTIQQDADWDSVLSSQQRMESENNKLCSLYSFRNTSTSPHKPDEGSRDREIMTSVTF
SEG     ............................................................
COILS   ............................................................
1nhm-   ............................................................

SEQ     GTPERRKGSLADVVDTLKQKKLEEMTRTEQEDSSCMEKLLSKDWKEKMERLNTSELLGEI
SEG     ............................................................
COILS   ............................................................
1nhm-   ............................................................

SEQ     KGTPESLAEKERQLSTMITQLISLREQLLAAHDEQKKLAASQIEKQRQQMDLARQQQEQI
SEG     .............................................xxxxxxxxxxxxxxx
COILS   ......................................................CCCCCC
1nhm-   ............................................................

SEQ     ARQQQQLLQQQHKINLLQQ[illegible]GHMPPLMIPIFPHDQRTLAAAAAAQQGFLFPPGITY
SEG     xxxxxxxxxxxxxxxxxxxxxx....................xxxxxx............
COILS   CCCCCCCCCCCCCCCCCCCCCC......................................
1nhm-   ............................................................

SEQ     KPGDN[illegible]VQFIPSTMAAAAASGLSPLQLQQLYAAQLASMQVSPGAKMPSTPQPPNTAGTV
SEG     .........................xxxxxxxxxxxx.......................
```

```
COILS  ............................................................
1nhm-  ............................................................

SEQ    SPTGIKNEKRGTSPVTQVKDEAAAQPLNLSSRPKTAEPVKSPTSPTQNLFPASKTSPVNL
SEG    ............................................................
COILS  ............................................................
1nhm-  ............................................................

SEQ    PNKSSIPSPIGGSLGRGSSLGKWKSQHQEETYELDILSSLNSPALFGDQDTVMKAIQEAR
SEG    ...xxxxxxxxxxxxxxxxxx.......................................
COILS  ............................................................
1nhm-  ............................................................

SEQ    KMREQIQREQQQQPHGVDGKLSSINNMGLNSCRNEKERTRFENLGPQLTGKSNEDGKLG
SEG    ..xxxxxxxxxxxx..............................................
COILS  ............................................................
1nhm-  ............................................................

SEQ    PGVIDLTRPEDAEGSKAMNGSAAKLQQYYCWPTGGATVAEARVYRDARGRASSEPHIKRP
SEG    ............................................................
COILS  ............................................................
1nhm-  .........................................................CCC

SEQ    MNAFMVWAKDERRKILQAFPDMHNSNISKILGSRWKSMSNQEKQPYYEEQARLSKIHLEK
SEG    ...........................................................x
COILS  ............................................................
1nhm-  CCCHHHHHHHHHHHHHHHTTTTCCHHHHHHHHHHHHTTTTTTHHHHHHHHHHHHHHHHHH

SEQ    YPNYKYKPRPKRTCIVDGKKLRIGEYKQLMRSRRQEMRQFFTVGQQPQIPITTGTGVVYP
SEG    xxxxxxxxxxxx................................................
COILS  ............................................................
1nhm-  HHHTTTTTTT..................................................

SEQ    GAITMATTTPSPQMTSDCSSTSASPEPSLPVIQSTYGMKTDGGSLAGNEMINGEDEMEMY
SEG    .....................................................xxxxxxx
COILS  ............................................................
1nhm-  ............................................................

SEQ    DDYEDDPKSDYSSENEAPEAVSAN
SEG    xxxxxx..................
COILS  ........................
1nhm-  ........................
```

Prosite for DKFZphtes3_17n12.1

PS00001	97->101	ASN_GLYCOSYLATION	PDOC00001
PS00001	172->176	ASN_GLYCOSYLATION	PDOC00001
PS00001	388->392	ASN_GLYCOSYLATION	PDOC00001
PS00001	422->426	ASN_GLYCOSYLATION	PDOC00001
PS00001	559->563	ASN_GLYCOSYLATION	PDOC00001
PS00001	626->630	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	369->373	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	28->31	PKC_PHOSPHO_SITE	PDOC00005
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	136->139	PKC_PHOSPHO_SITE	PDOC00005
PS00005	203->206	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	390->393	PKC_PHOSPHO_SITE	PDOC00005
PS00005	512->515	PKC_PHOSPHO_SITE	PDOC00005
PS00005	530->533	PKC_PHOSPHO_SITE	PDOC00005
PS00005	692->695	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00006	547->551	CK2_PHOSPHO_SITE	PDOC00006
PS00006	577->581	CK2_PHOSPHO_SITE	PDOC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDOC00006
PS00006	793->797	CK2_PHOSPHO_SITE	PDOC00006
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	431->437	MYRISTYL	PDOC00008

PS00008	437->443	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	575->581	MYRISTYL	PDOC00008
PS00008	762->768	MYRISTYL	PDOC00008
PS00009	677->681	AMIDATION	PDOC00009
PS00017	526->534	ATP_GTP_A	PDOC00017
PS00029	187->209	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphtes3_17n12.1

```
HMM_NAME       HMG (high mobility group) box

HMM                *PKRPMNAYMLWMQEMRekIKaENPNdMhNtEISKMiGEMWKnMsEEEKm
                    +KRPMNA+M+W+++ R+KI +   P DMHN++ISK++G +WK+MS +EK+
Query          597  IKRPMNAFMVWAKDERRKILQAFP-DMHNSNISKILGSRWKSMSNQEKQ    644

HMM                PYEdMAeeEKqRYMKEMPeYK*
                   PY+++ +++ + +++ +P+YK
Query          645 PYYEEQARLSKIHLEKYPNYK    665
```

DKFZphtes3_17n18
----

group: intracellular transport and trafficking

DKFZphtes3_17n18 encodes a novel 782 amino acid protein with weak partial similarity to known proteins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a TonB-dependent receptor protein signature 1. In E. coli, the tonB protein interacts with outer membrane receptor proteins that mediate uptake of specific substrates into the periplasmic space. In the absence of tonB these receptors bind their substrates but do not carry out active transport. The novel protein seems to be involved in ATP-dependent transport of substances into the cell.

The new protein can find application in modulation of cell-permeability and transport of suitable substrates into the cell.

unknown receptor

protein containes TONB_DEPENDENT_REC_1 Pattern and ATP_GTP_A Pattern,

Sequenced by GBF

Locus: unknown

Insert length: 2853 bp
Poly A stretch at pos. 2806, no polyadenylation signal found

```
   1 GTCCTTTTAA GTCAGTAAAT TGAACTAAGT CGGTTATTCG GCAAGCAGTT
  51 CCTATAAAAA ACTACATGGC TAAGGTTCTT AATGATTGAC CACAAGCAGA
 101 TCTTTCACCC TCGGATCTCT AGCTACAAAA GGTCCCCACA CTGAAGAAGC
 151 CACTACCTCC ACCACCACCA GCACCACCAC GTCCAGTGCT GCTGGCAACC
 201 ACTGGGGCAG CCAAGCGCTC CACCCTCTCT CCCACCATGG CCCGTCAGGT
 251 GCGCACCCAC CAGGAGACCC TGAACAGGTT TCAGCAGCAG TCCATCCACC
 301 TGCTGACGGA GCTCCTCAGA CTGAAGATGA AGGCCATGGT GGAGTCTATG
 351 TCGGTGGGTG CCAACCCCTT GGACATCACC AGGCGCTTTG TGGAGGCCAG
 401 CCAGCTCCTC CACCTCAATG CCAAGGAGAT GGCCTTCAAC TGCCTGATCA
 451 GCACAGCCGG GAGAAGTGGC TACAGCAGCG GACAGTTGTG GAAAGAGTCC
 501 CTCGCAAACA TGTCCGCCAT TGGGGTGAAC TCGCCTTACC AGCTGATCTA
 551 CCACTCTTCC ACAGCCTGTC TGAGCTTTTC TCTCTCTGCT GGAAAAGAAG
 601 CCAAGAAGAA AATAGGCAAA TCTAGAACTA CAGAAGATGT CAGCATGCCG
 651 CCCCTGCATC GAGGAGTGGG AACCCCTGCC AACAGCCTGG AGTTCAGCGA
 701 CCCCTGCCCT GAGGCCCGGG AGAAGCTGCA GGAGTTGTGT CGCCACATAG
 751 AAGCTGAAAG GGCCACATGG AAAGGGAGGA ATATCTCCTA CCCCATGATC
 801 TTACGAAACT ACAAGGCAAA GATGCCCTCT CATCTAATGT TGGCCCGCAA
 851 AGGAGACTCT CAGACCCCGG GTTTACATTA CCCTCCCACT GCAGGTGCTC
 901 AGACTCTCAG CCCCACCTCT CACCCATCTT CTGCCAACCA TCATTTCAGT
 951 CAGCATTGTC AAGAGGGGAA GGCACCCAAG AAGGCCTTCA AGTTTCATTA
1001 CACCTTCTAT GATGGCTCCT CCTTCGTTTA CTATCCCTCT GGAAACGTCG
1051 CTGTATGTCA GATCCCCACA TGCTGCAGAG GGAGAACCAT CACCTGCCTC
1101 TTTAATGACA TACCTGGATT CTCCTTGCTG GCCCTATTCA ATACTGAAGG
1151 CCAGGGCTGT GTTCACTACA ACCTAAAAAC CAGTTGCCCA TATGTCTTAA
1201 TCTTGGATGA GGAAGGTGGG ACCACCAATG ACCAGCAGGG CTATGTAGTC
1251 CACAAGTGGA GCTGGACTTC CAGGACAGAG ACCCTGCTTT CCCTGGAATA
1301 CAAGGTGAAT GAGGAAATGA AACTAAAGGT ACTGGGACAG GACTCCATCA
1351 CAGTCACCTT CACCTCCCTG AATGAGACAG TAACACTCAC TGTGTCGGCC
1401 AACAATTGTC CCCATGGAAT GGCATATGAC AAACGGCTGA ACCGCAGAAT
1451 CAGCAACATG GACGACAAGG TGTATAAGAT GAGCCGAGCC CTGGCTGAGA
1501 TCAAGAAGCG GTTTCAGAAG ACAGTGACTC AGTTCATTAA TTCTATCTTG
1551 CTGGCCGCAG GTCTGTTTAC CATTGAATAT CCCACCAAAA AGGAGGAGGA
1601 AGAATTTGTT CGGTTCAAGA TGAGATCCAG AACTCATCCC GAGCGGCTCC
1651 CCAAGCTAAG TTTATACTCA GGAGAAAGTC TTTTACGATC TCAGTCAGGC
1701 CACCTGGAAT CCTCAATTGC AGAGACTTTG AAGGATGAGC CTGAGTCTGC
1751 TCCTGTGAGC CCAGTTCGGA AGACCACCAA AATCCACACC AAAGCCAAGG
1801 TCACATCCAG AGGGAAGGCC CGCGAGGGGC GCAGCCCCAC CAGGTGGGCG
1851 GCCTTGCCCT CAGACTGCCC GCTGGTGCTG CGGAAGCTCA TGCTCAAGGA
1901 AGACACCCGT GCTGGCTGCA AGTGCCTGGT GAAGGCGCCC CTGGTCTCTG
1951 ACGTGGAGCT GGAGCGCTTC CTGTTGGCGC CCCGAGACCC CAGCCAAGTG
2001 CTGGTGTTTG GGATCATCTC AAGCCAGAAC TACACCAGCA CTGGGCAGCT
2051 CCAGTGGCTG CTGAACACTC TCTACAACCA CCAGCAGCGG GGCCGTGGCT
2101 CCCCCTGCAT CCAGTGCCGG TATGACTCCT ACCGCCTGCT GCAGTATGAC
2151 CTGGACAGCC CCCTGCAGGA GGACCCTCCC CTGATGGTGA AGAAGAACTC
2201 TGTGGTGCAG GGGATGATTC TGATGTTTGC CGGGGGGAAG CTCATTTTTG
2251 GGGGCCGTGT TTTGAATGGA TATGGCCTCA GCAAGCAGAA TCTGCTGAAA
2301 CAGATCTTCC GGTCTCAACA GGATTACAAG ATGGGCTACT TCCTGCCGGA
2351 TGACTACAAA TTCAGTGTTC CCAACTCTGT CCTGAGCCTG GAGGATTCTG
2401 AATCAGTCAA GAAAGCCGAG TCAGAAGATA TCCAAGGAAG CAGCTCCTCA
```

```
2451 TTGGCCCTGG AAGACTATGT GGAGAAGGAG TTATCTCTGG AGGCTGAGAA
2501 GACAAGAGAG CCTGAAGTGG AGCTACATCC TCTCAGCAGG GACAGCAAGA
2551 TAACTAGTTG GAAGAAGCAG GCCTCCAAGA AGTAGCGCCA TCCTGGCAGC
2601 AGCCAAGTGA GCCAGGCCCC GGCCCGGGGT GCTGGGGCTT CTTGCCAGCC
2651 CAGCCCTGCC TCCCCGGTCT CCCACCCTGT CCTCCAAGCT TCTATAATAA
2701 ACCAGCGGGC CTCCAGCATT GGGGTGAGGC TCTGGGGAAG GACAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAGGG
2801 CGGCCGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGGGCGG
2851 CCG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 237 bp to 2582 bp; peptide length: 782
Category: putative protein
Prosite motifs: ATP_GTP_A (122-130)
TONB_DEPENDENT_REC_1 (1-44)

```
  1 MARQVRTHQE TLNRFQQQSI HLLTELLRLK MKAMVESMSV GANPLDITRR
 51 FVEASQLLHL NAKEMAFNCL ISTAGRSGYS SGQLWKESLA NMSAIGVNSP
101 YQLIYHSSTA CLSFSLSAGK EAKKKIGKSR TTEDVSMPPL HRGVGTPANS
151 LEFSDPCPEA REKLQELCRH IEAERATWKG RNISYPMILR NYKAKMPSHL
201 MLARKGDSQT PGLHYPPTAG AQTLSPTSHP SSANHHFSQH CQEGKAPKKA
251 FKFHYTFYDG SSFVYYPSGN VAVCQIPTCC RGRTITCLFN DIPGFSLLAL
301 FNTEGQGCVH YNLKTSCPYV LILDEEGGTT NDQQGYVVHK WSWTSRTETL
351 LSLEYKVNEE MKLKVLGQDS ITVTFTSLNE TVTLTVSANN CPHGMAYDKR
401 LNRRISNMDD KVYKMSRALA EIKKRFQKTV TQFINSILLA AGLFTIEYPT
451 KKEEEEFVRF KMRSRTHPER LPKLSLYSGE SLLRSQSGHL ESSIAETLKD
501 EPESAPVSPV RKTTKIHTKA KVTSRGKARE GRSPTRWAAL PSDCPLVLRK
551 LMLKEDTRAG CKCLVKAPLV SDVELERFLL APRDPSQVLV FGIISSQNYT
601 STGQLQWLLN TLYNHQQRGR GSPCIQCRYD SYRLLQYDLD SPLQEDPPLM
651 VKKNSVVQGM ILMFAGGKLI FGGRVLNGYG LSKQNLLKQI FRSQQDYKMG
701 YFLPDDYKFS VPNSVLSLED SESVKKAESE DIQGSSSSLA LEDYVEKELS
751 LEAEKTREPE VELHPLSRDS KITSWKKQAS KK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_17n18, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_17n18, frame 3

Report for DKFZphtes3_17n18.3

[LENGTH]	782
[MW]	88030.16
[pI]	9.22
[BLOCKS]	BL00286 Squash family of serine protease inhibitors proteins
[PROSITE]	ATP_GTP_A 1
[PROSITE]	MYRISTYL 4
[PROSITE]	CAMP_PHOSPHO_SITE 3
[PROSITE]	CK2_PHOSPHO_SITE 14
[PROSITE]	PROKAR_LIPOPROTEIN 1
[PROSITE]	TONB_DEPENDENT_REC_1 1
[PROSITE]	PKC_PHOSPHO_SITE 10
[PROSITE]	ASN_GLYCOSYLATION 4
[KW]	Alpha_Beta

```
SEQ    MARQVRTHQETLNRFQQQSIHLLTELLRLKMKAMVESMSVGANPLDITRRFVEASQLLHL
PRD    cchhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhh

SEQ    NAKEMAFNCLISTAGRSGYSSGQLWKESLANMSAIGVNSPYQLIYHSSTACLSFSLSAGK
PRD    hhhhhhhhhhhcccccccccccchhhhhhhhhcccccccceeeecccceeeecccccccch

SEQ    EAKKKIGKSRTTEDVSMPPLHRGVGTPANSLEFSDPCPEAREKLQELCRHIEAERATWKG
PRD    hhhhhhccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhc

SEQ    RNISYPMILRNYKAKMPSHLMLARKGDSQTPGLHYPPTAGAQTLSPTSHPSSANHHFSQH
PRD    cccccchhhhhhhcccccceeecccccccccccccccccccccccccccccccccccccc

SEQ    CQEGKAPKKAFKFHYTFYDGSSFVYYPSGNVAVCQIPTCCRGRTITCLFNDIPGFSLLAL
PRD    ccccccchhhheeeecccccceeeecccceeeeeccccccceeeeeeccccccccceeeee

SEQ    FNTEGQGCVHYNLKTSCPYVLILDEEGGTTNDQQGYVVHKWSWTSRTETLLSLEYKVNEE
PRD    eccccccceeeeecccccccceeeeeccccccccccceeeeeeeecccchhhhhhhhhhhhhh

SEQ    MKLKVLGQDSITVTFTSLNETVTLTVSANNCPHGMAYDKRLNRRISNMDDKVYKMSRALA
PRD    hhhhhhccceeeeeeecccccceeeeeeeeccccccchhhhhhhhhhhcccchhhhhhhhh

SEQ    EIKKRFQKTVTQFINSILLAAGLFTIEYPTKKEEEEFVRFKMRSRTHPERLPKLSLYSGE
PRD    hhhhhhhhhhhhhhhhhhhhcccceeecccchhhhhhhhhhhcccccccccccceeeeccc

SEQ    SLLRSQSGHLESSIAETLKDEPESAPVSPVRKTTKIHTKAKVTSRGKAREGRSPTRWAAL
PRD    eeeecccccchhhhhhhhcccccccccccccccccccceeeeeccccccccccccccccccc

SEQ    PSDCPLVLRKLMLKEDTRAGCKCLVKAPLVSDVELERFLLAPRDPSQVLVFGIISSQNYT
PRD    ccccchhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhccccccceeeeeeeecccccc

SEQ    STGQLQWLLNTLYNHQQRGRGSPCIQCRYDSYRLLQYDLDSPLQEDPPLMVKKNSVVQGM
PRD    cchhhhhhhhhhhhhhcccccccceeeecccccceeecccccccccccccccccchhhhh

SEQ    ILMFAGGKLIFGGRVLNGYGLSKQNLLKQIFRSQQDYKMGYFLPDDYKFSVPNSVLSLED
PRD    heeeccccccccccccccccccchhhhhhhhhhhhhcccccccccccceeecccceeeccc

SEQ    SESVKKAESEDIQGSSSSLALEDYVEKELSLEAEKTREPEVELHPLSRDSKITSWKKQAS
PRD    chhhhhhhccccccccccchhhhhhhhhhhhhhhhhcccceeeccccccccccccccccc

SEQ    KK
PRD    cc
```

Prosite for DKFZphtes3_17n18.3

PS00001	91->95	ASN_GLYCOSYLATION	PDOC00001
PS00001	182->186	ASN_GLYCOSYLATION	PDOC00001
PS00001	379->383	ASN_GLYCOSYLATION	PDOC00001
PS00001	598->602	ASN_GLYCOSYLATION	PDOC00001
PS00004	403->407	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	511->515	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	652->656	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	177->180	PKC_PHOSPHO_SITE	PDOC00005
PS00005	344->347	PKC_PHOSPHO_SITE	PDOC00005
PS00005	450->453	PKC_PHOSPHO_SITE	PDOC00005
PS00005	497->500	PKC_PHOSPHO_SITE	PDOC00005
PS00005	513->516	PKC_PHOSPHO_SITE	PDOC00005
PS00005	523->526	PKC_PHOSPHO_SITE	PDOC00005
PS00005	631->634	PKC_PHOSPHO_SITE	PDOC00005
PS00005	723->726	PKC_PHOSPHO_SITE	PDOC00005
PS00005	774->777	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	131->135	CK2_PHOSPHO_SITE	PDOC00006
PS00006	256->260	CK2_PHOSPHO_SITE	PDOC00006
PS00006	329->333	CK2_PHOSPHO_SITE	PDOC00006
PS00006	345->349	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	406->410	CK2_PHOSPHO_SITE	PDOC00006
PS00006	450->454	CK2_PHOSPHO_SITE	PDOC00006
PS00006	466->470	CK2_PHOSPHO_SITE	PDOC00006
PS00006	493->497	CK2_PHOSPHO_SITE	PDOC00006
PS00006	497->501	CK2_PHOSPHO_SITE	PDOC00006
PS00006	571->575	CK2_PHOSPHO_SITE	PDOC00006
PS00006	693->697	CK2_PHOSPHO_SITE	PDOC00006
PS00006	717->721	CK2_PHOSPHO_SITE	PDOC00006
PS00008	145->151	MYRISTYL	PDOC00008
PS00008	327->333	MYRISTYL	PDOC00008
PS00008	592->598	MYRISTYL	PDOC00008
PS00008	734->740	MYRISTYL	PDOC00008

PS00013	101->112	PROKAR_LIPOPROTEIN	PDOC00013
PS00017	122->130	ATP_GTP_A	PDOC00017
PS00430	1->44	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKFZphtes3_17n18.3)

DKFZphtes3_18f3
----

group: testes derived

DKFZphtes3_18f3 encodes a novel 248 amino acid protein with partial similarity to human TNF-inducible protein CG12-1.

The novel protein contains two leucine zippers.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to TNF-inducible protein CG12-1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4608 bp
Poly A stretch at pos. 4570, polyadenylation signal at pos. 4550

```
   1 GACAGAAGTG AATGGGAATG GAGAGGCCGG CGGCCCGGGA GCCGCATGGG
  51 CCCGACGCGC TGCGGCGCTT CCAGGGACTG CTGCTGGACC GCCGAGGCCG
 101 GCTGCACCGC CAGGTGCTGC GCCTGCGCGA GGTGGCCCGG CGCCTGGAGC
 151 GCCTGCGCAG GCGCTCCCTC GTAGCCAACG TGGCCGGCAG CTCGCTGAGC
 201 GCAACGGGCG CCCTCGCCGC CATCGTGGGG CTCTCGCTCA GCCCGGTCAC
 251 CCTGGGGACC TCGCTGCTGG TGTCGGCCGT GGGGCTGGGG GTGGCCACAG
 301 CCGGAGGGGC CGTCACCATC ACGTCCGATC TCTCGCTGAT CTTCTGCAAC
 351 TCCCGGGAGC TGCGGAGGGT GCAGGAGATC GCGGCCACCT GCCAGGACCA
 401 GATGCGAGAG ATCCTGAGCT GCCTCGAGTT TTTCTGCCGC TGGCAGGGCT
 451 GCGGGGACCG CCAGCTGCTG CAGTGCGGGA GGAACGCCTC CATCGCCCTG
 501 TACAATTCTG TCTACTTCAT CGTCTTCTTT GGCTCACGTG GCTTCCTCAT
 551 CCCCAGGCGG GCGGAGGGGG ACACCAAGGT TAGCCAGGCC GTGCTGAAGG
 601 CCAAGATTCA GAAACTGGCC GAGAGCCTGG AGTCCTGCAC CGGGGCTCTG
 651 GACGAACTCA GCGAGCAGCT GGAGTCTCGG GTTCAGCTCT GCACCAAGTC
 701 CAGTCGTGGC CACGACCTCA AGATCTCTGC TGACCAGCGT GCAGGGCTGT
 751 TTTTCTGAGA ACATCCTTTC CCCCTAATGA CCGAGGCCAG CAAATCATCC
 801 TCATGGGATG CTCCAGAATT TGTAGCTCCC TTAGGAAAAC ACCAAGCTGG
 851 GTTAGGAGCC GAAGGCAAAG GATGAGAAAA ACTGTTTTTG AAGTGGGCAG
 901 GTCCCCAAAG CCCTTCTTTT CCCATCACTG TGACATCTGC CTGGGCTTGA
 951 GTGCTACGGA CTTTTCAGTC TTCCTAGTGG AAAAATGTGA CCCAAAAACT
1001 CCTTTTCCTT TATCAAAAAC TTTCTGTCTA AACACAGCTG GGCAGGCACT
1051 CCTGTTTTAA AGTTATTTCG GGGTCCCTGA CCCTGCCCTG GTGGCTTGGC
1101 CTGAGACTGG AGAGAGTGCC ATCCTCTGGG TCCTCTCCAA GTCCTACTAG
1151 TCTTTGAAGT CCTCAAAATG TGCGTGAGGA AGGCATTTGC CTCTATTCCA
1201 GAATTTCTGA TACAAAGAAC TCCAGAATCC AGAGCAAATC AGCCCTTCTC
1251 TGAACGTTGT AGGATGGTTC AGAACCCAGA GAGGACCCTG GTGCTGATAT
1301 CTCCTCCTCT TCCCTTTCCC CTCAGCTTAC TTACTCCCAG ATGCGGCCTG
1351 GGTATGAAGT AGGCCTTTCC TGAGTGGCTC CCAATCCAGT CCTCCAAGTA
1401 CTCAGAGGGG AAGCCCGTGA AGCCGTCATC TAAGTCCTGC TCCCTCACAT
1451 GAAGCTGAGG GCCAGATAGA TGGAGCGACT GCCAACTTCA TTTCCCGACA
1501 TCATTGTGTT CAGAAGAGAG TGATGGGTTT TGAGTTAGAC AGTCCTGGGC
1551 TTGAGACAGG CTTTGTCACT ACTGTGTGAG TGTAGCCACC TAATCTCTCT
1601 GAGACTGTGT AAAACAAAGA TGATAAAATC TCACCCTGTT GTGAGATATT
1651 AAATGAGCCA AAGTGCCTAG CATGATGGTG CTGGCTCATA TAGTGTAGTC
1701 CCTGGAATGG CAAATTAACA TCACCCAGGA ACTTGTTAGA AAGGCAAATT
1751 CTTGGACACA ACCCTCCTGA TTTATGGAAT CAGAAACTCT GGCTGTGGGG
1801 CCCAGCAACC TGAGTTTAAA CAATTTCTCT GGGTGGTTCT GCGGCACACT
1851 AAGGTTTGAA AATCACTACA ACAAATGCTA ACTTCTAATC CCCTTGATGA
1901 GCTTTCACGA AGTCTCACGG CTTCTCTAGG GACTCCATGG TCTTCAGAGT
1951 CGTTCACAGA TGACCAAGGA CAGACTGTGT CCCAGAAGCC AAAATGAGAG
2001 AGAGAGAGAG AGCACGCGTA CGTGCACCCT GGGGCAGTGT CTCACCGTAT
2051 GAATAAGGGA TGTAACACTA AAAGCCCATT AGGGGGCAGT GTTTCCCGCC
2101 TGTTGTAGAA ACTGGTACAG AAAGGATCCT ATATGAAGTT CCTGAAACTG
2151 ACCTTTGTCT ATTATTACCT TCTCTGAAAA GTGCCAGTCC ATGTATTTTT
2201 TATTTATTTT AAGTTTGTAA TTTAATTTTT AATTATTGTT TAGTGTTTGC
2251 ATTTAATTTT ATTTAATCAC CACATTTAGA AAATAATAAG AGCAAGTTTC
2301 TAAATGGGAG ACTGCTGAGG CTCTTTGCAA GAGATGAGAT TAAGTTTGAG
2351 TTTCTAAGGC AGGGCATGAG CTGGAAATAG CATTGCTTTC CTTGATTGTC
2401 TCTCTCCTTC AGGGAGATTC TTTTTCTCTA GTGTTTTAAG TGATCCTTTG
2451 AAGTAAGTGT GGAGAGTCTT GAATGGCAAG ACCAGGAGCT GAGTTTAAGC
2501 TTGTAATGGA AGCTTGCATT GTGGGATATA TAACTGAGGA AGCATATTTA
2551 TCCTGAAGGT ATTTTGCCAG AAGGTATCAC TTGACCTGGA AAAGGAATCT
2601 ATTTAGTTCA GGAAAGATAA AAAGTTTAGA GGTATGTGAA GGAAGCACTT
2651 AGAACTTGCA AGCCTGATGT CCTATCAAGT TATGTCTTCT GGGTGACAGA
2701 CAAAATAGCT TGTCTTATGG TGGTGATGTG TTGCATTTTC ACTTTGGGGT
```

```
2751 CTGTAAGAAA CTGTCAGTGA AAATATGTAC AATTCCTTCA ATTTCCATTC
2801 TTAACAACTG TAATGTTGAA AAATAAGTTG AAAAGTCTTT GGGACCATAC
2851 ATGCAAAAAC GGTGCCTCTG TTACTTAATT ATTTAATATT CTATAAATGT
2901 ACCCAATCTG TCCGCACCCT TCCCAGTGAT GGGGCAGTAT GTCTGAGGAA
2951 GTATAATTTC AGTACTGGGG TCGGGGAGAG GAGGTGATGT TTCTACATTT
3001 TTATTTTTTC TATAAATTGC AATTGGTCTG TATGCTGGTT TATTTTGAAA
3051 TTTATATTGG TTTCTTTTCA AGCTGGTGTC ATCTCCTAGA CTGTTTCACC
3101 CAGATGCTAG CATTTTTTTT TTTTTTGAGA CAGAGTCTCA CTCTGTCACC
3151 TAGGCTGGAG TTGCAGTGGT TTGATCTCGG CTCACTGCAA CCTCCGACTC
3201 CTGGGTTCAA GCAATTCTTC TGCCTCAGCC TCCTGAGTAG CTGGGATTAC
3251 AGATGTGCAC CAGCACACCC GGCTAATTTT TTGTATTTTT AGTAGAGACA
3301 GGGTTTCGCC ATGTTGGCCA GGCTGGTCTT GAACTCCTGG CCTTATGTGA
3351 TCCGCCCACC TTGGCTTCCC AAAGTGCTGG GATTACAGGC ATGAGCCACC
3401 TCGCCTGGCC AGATGCTAGC ATTTTAGATC AAACAATTCA TTTTAGATGA
3451 ATTGTTTTGT TTCACAATCA TTTTAAATCA TTTTAGAATG TACTTCACAT
3501 TATTAGTTGT GTTATGGCAT AAAGGTACAA CCATTCCCTA ACTCCATCTT
3551 TTATTAATGC TTAAGTTTAA ATTATATTCT TCCAATGCCT AAGCTATTCC
3601 CTAGAATTAA ACTGGGCACT TTTGGAAGCA GCAACAGTAA CAGCAGCAGC
3651 AAACTTTTCC TCTCATATTT TGGGTGTATC AAAAGTTCTA GACTTTTGAA
3701 GTTATGATTT CAGTGGCCCA CTTTATTTCT AAGGAAGAGT GTCTACTTTG
3751 GAACGATACT TTGCACATAG TAGGAACTCA AGAAATACAT TTGAATAATT
3801 ATAATTAACT GTTTAGCTAT CTTAATGAGA ATTTGTTGAC AACAAAAGAT
3851 CATCCATCGC CTTATGTGTG AGTAAGATTG GAGCCTCTAT CAAGATTTAG
3901 TCAAGTTCAG TTAGATTGAT TCTAGAAACA AATATTTATT TCTTTCTTTT
3951 ACGGGGATGT GAATAAGGCT TTTCCTTAAG GCCTTCATTC TTTAAACAAA
4001 CAGGTTGAAA TGGTATGTTG TAAAAGAGAA GACGGGAGAG AGGTATTTAG
4051 ATGATAAGTG TACTTCACAA AAATGCCAAA GTTTGAAAAA TAGGTATGTT
4101 TGTTCTAAAT GTTTAAGTGC TTCTCTGTTA GGTTCTGGGG CTTGCAATCA
4151 TTTGAATTGT TCTGTTTCAC AATAAAGGAG ATTCACTGGG TTCTGCATTT
4201 TCAGGATTCA ATAGAACTGC TCCATTAAAA AAATAATCCT TAGCAAGCAT
4251 TCGAATCCTA ACTGCTTTGA TGCACTTGCC CTCGGGCACC TGTCATTTCC
4301 AATATGGTAG GTGTCAAAGT CAAAAGTATT TACTGGGAGA AAAAAGAGAG
4351 GAGTGGTTGT AGAAGTCTCC CTAAATCAGA CATGTCAAGC AATCAGCCAA
4401 CGTGGTGTAT TTCTCATTCA ATATTTTAGT GTGAATTGAG ACACTGAGAT
4451 AAAGACATCG TGCAGAGATA AATGGGGATA CAGTTAAATG TAGCAACTCT
4501 TGAGTTCATT TTTTCCCACT GTAGCAAAAT TAATGCTTTC TCTTTATTGA
4551 AATAAATTGC TCATTCCTCC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
4601 AAAAAAGG
```

BLAST Results

Entry HSG27587 from database EMBL:
human STS SHGC-32548.
Score = 1951, P = 9.0e-101, identities = 411/425

Entry HS073350 from database EMBL:
human STS EST303564.
Score = 1417, P = 8.7e-58, identities = 285/287

Medline entries

No Medline entry

Peptide information for frame 2

ORF from the beginning to 580 bp; peptide length: 194
Category: questionable ORF
Classification: no clue

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_18f3, frame 2

PIR:CGBO1S collagen alpha 1(I) chain - bovine (fragments), N = 1, Score = 155, P = 4.5e-10

TREMBL:HSCG1PA1_1 gene: "COL1A1"; Human proalpha 1 (I) chain of type I procollagen mRNA (partial)., N = 1, Score = 155, P = 6.5e-10

```
>PIR:CGBO1S collagen alpha 1(I) chain - bovine (fragments)
            Length = 779

  HSPs:

 Score = 155 (23.3 bits), Expect = 4.5e-10, P = 4.5e-10
 Identities = 60/152 (39%), Positives = 67/152 (44%)

Query:      7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGA--APARGGPAPGAPAQALPRSQRGR 62
              G+ G PG + AR     PG      GPP PA P GA  AP   G     A A    P SQ
Sbjct:    230 GDLGAPGPSGARGERGFPGERGVEGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQGAP 289

Query:     63 QLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAAE 122
                L     G P    RGA    PG  GD       +GA G    + G      VR L    +  PG A
Sbjct:    290 GL---QGMPGE-RGAAGLPGPKGDRGDAGPKGADGAPGKDG----VRGLTGPIGPPGPAG 341

Query:    123 GAGDRGHL-P-GP--------DARDPELPRVFLPLAGLRGPPAA 156
               GD+G    P GP          D  +P  P      P AG   GPP A
Sbjct:    342 APGDKGEAGPSGPAGTRGAPGDRGEPGPPG---P-AGFAGPPGA 381

 Score = 121 (18.2 bits), Expect = 5.4e-05, P = 5.4e-05
 Identities = 52/154 (33%), Positives = 60/154 (38%)

Query:      7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARG-----GPAPGAPAQALPRSQRG 61
              G  G PGAA  R      P    AGPP P  P G   ++G       GPA G P +  P     G
Sbjct:    434 GATGFPGAA-GRVGPPGPSGNAGPPGPPGPAGKEGSKGPRGETGPA-GRPGEVGPPGPPG 491

Query:     62 RQLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAA 121
                 A   G P     G       PG PG           RG  G    +RG        R    L    PG +
Sbjct:    492 P--AGEKGAPGAD-GPAGAPGTPGPQGIAGQRGVVGLPGQRGE----RGFPGL---PGPS 541

Query:    122 EGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAAVRE 160
                G +G            R P  P     +   GL GPP     + RE
Sbjct:    542 GEPGKQGPSGASGERGPPGP---MGPPGLAGPPGESGRE 577

 Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04
 Identities = 52/148 (35%), Positives = 62/148 (41%)

Query:      7 GEAGGPGAAWARRAAALPGTAAGPPRPAA---PPGAAPARGGPAPGAPAQALPRSQRG-R 62
              G  G PG     AR +A  PG A G P  A       PPG +    GP PG P   A      +G R
Sbjct:    416 GNVGAPGPKGARGSAGPPG-ATGFPGAAGRVGPPGPS-GNAGP-PGPPGPAGKEGSKGPR 472

Query:     63 QLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRH--HHVRSLADLLQLPGA 120
                      GRP      G +  PG PG              GA G       G         +          ++ LPG
Sbjct:    473 GETGPAGRP----GEVGPPGPPGPAGEKGAPGADGPAGAPGTPGPQGIAGQRGVVGLPGQ 528

Query:    121 AEGAGDRGH--LPGPDARDPEL-PRVFLPLAGLRGPP 154
                 G+RG    LPGP        +    P          +G RGPP
Sbjct:    529 R---GERGFPGLPGPSGEPGKQGPS---GASGERGPP 559

 Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04
 Identities = 54/162 (33%), Positives = 64/162 (39%)

Query:      7 GEAGGPGAAWARRAAALPGT--AAGPPRPAAPPGAAPARG--GPA--PGAPAQALPRSQR 60
              G  G PG     +      PG     A+GP  P  PPG      G  G A   PG P +   P     +
Sbjct:     29 GPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGRPGERGPPGPQ 88

Query:     61 G-RQLAERNGRP--RRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHV--RSLADLL 115
              G R L     G P   + HRG       G  GD          +G  G        G        +  R L
Sbjct:     89 GARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGPRGLPGFP 148

Query:    116 QLPGAA--EG-AGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
                   GAA    G AG+RG +PGP              P        AG +GPP   A
Sbjct:    149 GPKGAAGEPGKAGERG-VPGPPGAVG--PAGKDGEAGAQGPPGPA 190

 Score = 113 (17.0 bits), Expect = 5.4e-04, P = 5.4e-04
 Identities = 54/148 (36%), Positives = 58/148 (39%)

Query:      7 GEAGGPGAAWARRAAALPGTA-----AGPPRPAAP---PGAAPARGGPAP-GAPAQALPR 57
              G AG PGA      A    PG A         AGPP PA P    PG        G P P GA     A P
Sbjct:    374 GFAGPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPAGPPGPIGNVGAPGPKGARGSAGPP 433

Query:     58 SQRGRQLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQL 117
                G    A      P     G          PG PG          +G  G        GR    V
Sbjct:    434 GATGFPGAAGRVGPPGPSGNAGPPGPPGPAGKEGSKGPRGETGPAGRPGEVGP------- 486

Query:    118 PGAAEGAGDRGHLPGPD--ARDPELPRVFLPLAGLRG 152
              PG     AG++G  PG D   A   P   P       +AG RG
Sbjct:    487 PGPPGPAGEKG-APGADGPAGAPGTPGP-QGIAGQRG 521

 Score = 110 (16.5 bits), Expect = 1.3e-03, P = 1.2e-03
```

```
 Identities = 54/151 (35%), Positives = 60/151 (39%)

Query:     7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPG--AAPAR-GGPAP-GAPAQALPRSQRGR 62
             GE G  G A +      LPG A GPP  A  PG    P   G P P GA  +     +RG
Sbjct:  194 GERGEQGPAGSPGFQGLPGPA-GPPGEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGV 252

Query:    63 QLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAAE 122
             +       PR   GA    G GD  A    G+ G     +G      R A L    PG
Sbjct:  253 EGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQGAPGLQGMPGE-RGAAGL---PGPK- 307

Query:   123 GAGDRGHLPGPDARD--PELPRVFLPLAGLRGPPAAA 157
               GDRG   GP   D  P    V   L G  GPP  A
Sbjct:  308 --GDRGDA-GPKGADGAPGKDGV-RGLTGPIGPPGPA 340

 Score = 109 (16.4 bits), Expect = 1.7e-03, P = 1.7e-03
 Identities = 55/154 (35%), Positives = 60/154 (38%)

Query:     4 NGN-GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARG-GPAPGAPAQALPRSQRG 61
             NG+ GEAG PG        R    P  A G P  A  PG     RG      GA   A P  +G
Sbjct:   67 NGDDGEAGKPGRP-GERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKG 125

Query:    62 RQLAE-RNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSL-----ADLL 115
               +    NG P +  G     PG PG     A     G     G      V      A
Sbjct:  126 EPGSPGENGAPGQ-MGPRGLPGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQ 184

Query:   116 QLPGAAEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
               PG A   AG+RG   GP A   P        F  L G  GPP  A
Sbjct:  185 GPPGPAGPAGERGE-QGP-AGSPG----FQGLPGPAGPPGEA 220

 Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 44/131 (33%), Positives = 49/131 (37%)

Query:     2 EVNGNGEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAP-GAPAQALPRSQR 60
             E      GE G PG     R      LPG    GP    A  PG A  RG P P GA    A     +
Sbjct:  126 EPGSPGENGAPGQMGPR---GLPGFP-GPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEA 181

Query:    61 GRQLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGA 120
             G Q           P    RG       G PG     G+   G   G        G+          DL    PG
Sbjct:  182 GAQGPPGPAGPAGERGEQGPAGSPG--FQGLP-GPAGPPGEAGKPGEQGVPGDL-GAPGP 237

Query:   121 AEGAGDRGHLPG 132
             +    G+RG   PG
Sbjct:  238 SGARGERG-FPG 248

 Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 43/131 (32%), Positives = 55/131 (41%)

Query:     7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRQLAE 66
             GEAG   G A  R A    PG    G P P   P G A     GP PGA   Q    + +  G     A+
Sbjct:  347 GEAGPSGPAGTRGA---PGDR-GEPGPPGPAGFA----GP-PGADGQPGAKGEPGDAGAK 397

Query:    67 RNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAAEGAGD 126
             +   P     G       PG   G++ A     +GA G      G            + A   +   PG  +   AG
Sbjct:  398 GDAGPPGPAGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGA-AGRVGPPGPSGNAGP 456

Query:   127 RGHLPGPDARD 137
             G  PGP  ++
Sbjct:  457 PGP-PGPAGKE 466

 Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03
 Identities = 56/162 (34%), Positives = 62/162 (38%)

Query:     7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRQL 64
             G  G  PGA         A    G      GP   P   P  G  A  ARG      P   P  Q   PR   +G
Sbjct:  608 GPPGAPGAPGPVGPAGKSGDRGETGPAGPIGPVGPAGARG---PAGP-QG-PRGBKGZTG 662

Query:    65 AERNGRPRRHRG---ALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLA-DLLQ-LPG 119
             + +     + HRG              PG PG          GA G       RG          S     D L      LPG
Sbjct:  663 ZZGBRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGSPGKDGLNGLPG 722

Query:   120 AAEGAGDRGHL--PGPDARDPELPRVFLPLAGLRGPPAAAVREERLHRPVQ 168
                  G RG         GP A   P   P        P   G   GPP+            L  +P Q
Sbjct:  723 PIGPPGPRGRTGDAGP-AGPPGPPG---P-PGPPGPPSGGYDLSFLPQPPQ 768

 Score = 101 (15.2 bits), Expect = 1.5e-02, P = 1.5e-02
 Identities = 49/148 (33%), Positives = 55/148 (37%)

Query:     7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPA----QALPRSQRGR 62
             G AG PG A  R      PG A GP    A    G A A+G P P   PA      +   P    G
Sbjct:  152 GAAGEPGKAGERGVPGPPG-AVGP---AGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGF 207
```

```
Query:   63 QLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAAE 122
            Q      P    G   + G PGDL A    G  G     RG     R +      PG A
Sbjct:  208 QGLPGPAGPPGEAGKPGEQGVPGDLGAP---GPSGARGERGFPGE-RGVEGP---PGPAG 260

Query:  123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPP 154
             G  G  PG D     +       P  G +G P
Sbjct:  261 PRGANG-APGNDGAKGDAGAPGAP--GSQGAP 289

 Score = 100 (15.0 bits), Expect = 1.9e-02, P = 1.9e-02
 Identities = 40/130 (30%), Positives = 48/130 (36%)

Query:    7 GEAGGPGAAWARRAAALPGT--AAGPPRPAAPPGAAPARG--GPA--PGAPAQALPRSQR 60
            G G PG      +      PG    A+GP  P  PPG     G G A  PG P +  P   +
Sbjct:   29 GPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGRPGERGPPGPQ 88

Query:   61 G-RQLAERNGRP--RRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQL 117
            G R L     G P  + HRG     G  GD       +G  G     G    +     L
Sbjct:   89 GARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGPRG-LPGF 147

Query:  118 PGAAEGAGDRG 128
            PG     AG+ G
Sbjct:  148 PGPKGAAGEPG 158

 Score = 99 (14.9 bits), Expect = 2.5e-02, P = 2.5e-02
 Identities = 53/156 (33%), Positives = 61/156 (39%)

Query:    7 GEAGGPGAAWARRA---AALPGT--AAGPPRPAAPPGAAPARG--GPA----PGAPAQAL 55
            G  G PGA    R        A  PG    A G P P  P G +  RG  GPA    P  PA A
Sbjct:  587 GRDGSPGAKGDRGETGPAGAPGPPGAPGAPGPVGPAGKSGDRGETGPAGPIGPVGPAGAR 646

Query:   56 -------PRSQRGRQLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHV 108
                   PR  +G     + +    + HRG     G PG   +   +G G     G
Sbjct:  647 GPAGPQGPRGBKGZTGZZGBRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGP- 705

Query:  109 RSLADLLQLPGAAEGAGDRG--HLPGPDARDPELPRVFLPLAGLRGPP 154
                     PG+A    G  G    LPGP    P    PR    AG  GPP
Sbjct:  706 ---------PGSAGSPGKDGLNGLPGPIG--PPGPRGRTGDAGPAGPP 742

 Score = 98 (14.7 bits), Expect = 3.3e-02, P = 3.3e-02
 Identities = 51/158 (32%), Positives = 58/158 (36%)

Query:    7 GEAGGPGAAWARRAAALPGTA-----AGPPRPAAPPGAAPARGGPAP-GAPAQALPRSQR 60
            G  G  G     R AA LPG           AGP     PG     RG   P G P  A    +
Sbjct:  287 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGAPGKDGVRGLTGPIGPPGPAGAPGDK 346

Query:   61 GRQLAERNGRPRRHRGA---LAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQL 117
            G   A   +G P     RGA     +PG PG        GA G     +G     +   D
Sbjct:  347 GE--AGPSG-PAGTRGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGP- 402

Query:  118 PGAAEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAAVR 159
            PG A  AG  G +     A   P+  R        G  G P AA R
Sbjct:  403 PGPAGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGAAGR 444

 Score = 96 (14.4 bits), Expect = 5.7e-02, P = 5.5e-02
 Identities = 46/152 (30%), Positives = 57/152 (37%)

Query:    6 NGEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGAAPARGGPAPGAPA-QALPRSQRGR 62
            +G  G PGA  +         PG     G    PA PG   A G P P   PA ++  R + G
Sbjct:  574 SGREGAPGAEGSPGRDGSPGAKGDRGETGPAGAPGPPGAPGAPGPVGPAGKSGDRGETGP 633

Query:   63 QLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAAE 122
                   P     RG     G  G+       +G  G     RG     H  R  + L     PG
Sbjct:  634 AGPIGPVGPAGARGPAGPQGPRGB------KGZTGZZGBRGIKGH-RGFSGLQGPPGPPG 686

Query:  123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157
             G++G   P    A   P                AG RGPP +A
Sbjct:  687 SPGEQG--PS-GASGP---------AGPRGPPGSA 709

 Score = 94 (14.1 bits), Expect = 9.7e-02, P = 9.2e-02
 Identities = 45/134 (33%), Positives = 56/134 (41%)

Query:   24 PGTAAGPPRPAAPPGAAPARGGPA-PGAPAQALPRSQRGRQLAERNGRPRRHR--GALAQ 80
            P     G P P     PG     +G P   PG P   + P     RG         G P ++    G   +
Sbjct:   21 PSGPRGLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPP-----GPPGKNGDDGEAGK 75

Query:   81 PGHPGDLAA-GV--GRGAGGGHSRRGRHHHVRSLADLLQLPGAAEGAGDRGH--LPGPDA 135
            PG PG+    G      RG  G     G    H  R  + L    G  A  AG +G    PG +
Sbjct:   76 PGRPGERGPPGPQGARGLPGTAGLPGMKGH-RGFSGLDGAKGDAGPAGPKGEPGSPGENG 134

Query:  136 RDPEL-PRVFLPLAGLRGPPAAA 157
                ++  PR    LP    G    GP    AA
```

```
Sbjct:   135 APGQMGPRG-LP--GFPGPKGAA 154

 Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01
 Identities = 52/155 (33%), Positives = 58/155 (37%)

Query:     7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGP-APGAPAQALPRSQRGRQLA 65
             GEAG  G A  R A       G   GPP PA    G   A G P A G P  A  +   G
Sbjct:   347 GEAGPSGPAGTRGAPGDRGEP-GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGP 405

Query:    66 ERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGR--HHHVRSLADLLQLPGAA-- 121
                  P    G +  PG  G    +     GA G     GR          A     PG A
Sbjct:   406 AGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPAGK 465

Query:   122 EGA-GDRGHLPGPDARDPELPRVFLP-LAGLRGPPAA 156
             EG+ G RG    GP  R  E+      P AG +G P A
Sbjct:   466 EGSKGPRGET-GPAGRPGEVGPPGPPGPAGEKGAPGA 501

 Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01
 Identities = 51/156 (32%), Positives = 57/156 (36%)

Query:     7 GEAGGPGAAWARRA---AALPGT--AAGPPRPAAPPGAAPARGGPAPGAPAQAL-PRSQR 60
             G G PGA   R      A  PG   A G P P  P G +  RG    P  P   + P  R
Sbjct:   587 GRDGSPGAKGDRGETGPAGAPGPPGAPGAPGPVGPAGKSGDRGETGPAGPIGPVGPAGAR 646

Query:    61 GRQLAERNGRPRRHRGALAQPGHPGDLA-AGVG--RGAGGGHSRRGRH--HHVRSLADLL 115
             G   A   G PR  +G   +  G  G      G   +G  G     G         A
Sbjct:   647 GP--AGPQG-PRGBKGZTGZZGBRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPR 703

Query:   116 QLPGAAEGAGDRG--HLPGPDARDPELPRVFLPLAGLRGPP 154
              PG+A    G  G    LPGP    P  PR    AG  GPP
Sbjct:   704 GPPGSAGSPGKDGLNGLPGPIG--PPGPRGRTGDAGPAGPP 742

 Score = 90 (13.5 bits), Expect = 2.8e-01, P = 2.5e-01
 Identities = 45/134 (33%), Positives = 53/134 (39%)

Query:     7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRQ-LA 65
             G  G PG A  + A     G A  P  P  P G A  RG    G P Q   R +RG   L
Sbjct:   485 GPPGPPGPAGEKGAPGADGPAGAPGTPG-PQGIAGQRG--VVGLPGQ---RGERGFPGLP 538

Query:    66 ERNGRPRRH--RGALAQPGHPGDLA----AGV----GR-GAGGGHSRRGRHHHVRSLADL 114
             +G P  +    GA  + G PG +      AG    GR GA G     GR        + D
Sbjct:   539 GPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGAEGSPGRDGSPGAKGDR 598

Query:   115 LQL-PGAAEGAGDRGHLPGP 133
              +  P  A G        PGP
Sbjct:   599 GETGPAGAPGPPGAPGAPGP 618

 Score = 83 (12.5 bits), Expect = 1.8e+00, P = 8.3e-01
 Identities = 49/156 (31%), Positives = 56/156 (35%)

Query:     7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGAAPARG--GPAP--GAPAQALPRSQR 60
             G+AG  GA  A        + G    GPP PA  PG       G  GPA   GAP      R +
Sbjct:   311 GDAGPKGADGAPGKDGVRGLTGPIGPPGPAGAPGDKGEAGPSGPAGTRGAPGD---RGEP 367

Query:    61 GRQLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGA 120
             G           P      G   G PGD  A      G   G      G     +  ++    PG
Sbjct:   368 GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPAGPPGPIGNVG----APGP 423

Query:   121 AEGAGDRGHLPGPDARDPELPRVFLP----LAGLRGPPAAAVRE 160
                G  G  PG            RV  P    AG  GPP  A +E
Sbjct:   424 KGARGSAGP-PGATGFPGAAGRVGPPGPSGNAGPPGPPGPAGKE 466

 Score = 82 (12.3 bits), Expect = 2.3e+00, P = 9.0e-01
 Identities = 46/148 (31%), Positives = 52/148 (35%)

Query:     7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRQLAE 66
             G+AG PGA  ++ A  L  G    G      A  PG      RG  P  A      P     R L
Sbjct:   275 GDAGAPGAPGSQGAPGLQGMP-GERGAAGLPGPKGDRGDAGPKG-ADGAPGKDGVRGLTG 332

Query:    67 RNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAAEGAGD 126
                G P    G    PG  G+      G  G     RG          A       PGA     G
Sbjct:   333 PIGPP----GPAGAPGDKGEAGPSGPAGTRGAPGDRGEPGPPGP-AGFAGPPGADGQPGA 387

Query:   127 RGHLPGP-DARDPELPRVFLPLAGLRGPP 154
             +G  PG   A+     P    P AG  GPP
Sbjct:   388 KGE-PGDAGAKGDAGPPG--P-AGPAGPP 412
```

Peptide information for frame 3

```
ORF from 12 bp to 755 bp; peptide length: 248
Category: similarity to known protein
Classification: unset
Prosite motifs: LEUCINE_ZIPPER (17-39)
LEUCINE_ZIPPER (24-46)
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_18f3, frame 3

TREMBL:AF070675_1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds., N = 1, Score = 135, P = 1e-06

TREMBL:HS6802_1 gene: "dJ6802.1"; product: "dJ6802.1"; Homo sapiens DNA sequence from PAC 6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain, ESTs, CA repeat, STS and GSS., N = 1, Score = 107, P = 0.0023

```
>TREMBL:AF070675_1 product: "TNF-inducible protein CG12-1"; Homo sapiens
     TNF-inducible protein CG12-1 mRNA, complete cds.
          Length = 331

  HSPs:

 Score = 135 (20.3 bits), Expect = 1.0e-06, P = 1.0e-06
 Identities = 30/103 (29%), Positives = 55/103 (53%)

Query:    30 RLHRQVLRLREVARRLERLRRRSLVANVAGSSLSATGALAAIVGLSLSPVTLGTSLLVSA 89
             ++    + +LR +A  +E + R   ++NV  SS  A   + ++ GL L+P T GTSL ++A
Sbjct:    91 KIQESIEKLRALANGIEEVHRGCTISNVVSSSTGAASGIMSLAGLVLAPFTAGTSLALTA 150

Query:    90 VGLGVATAGGAVTITSDL-SLIFCNSRELRRVQEIAATCQDQMR 132
             G+G+  A     IT+ +    + +S E       + AT  D+++
Sbjct:   151 AGVGLGAASAVTGITTSIVEHSYTSSAEAE-ASRLTATSIDRLK 193
```

Pedant information for DKFZphtes3_18f3, frame 2

Report for DKFZphtes3_18f3.2

```
[LENGTH]      193
[MW]          19708.24
[pI]          11.90
[KW]          All_Alpha
[KW]          LOW_COMPLEXITY     55.44 %

SEQ    TEVNGNGEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQR
SEG    .......xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx...
PRD    cccccccccccccchhhhhhhhhccccccccccccccccccccccccccccccchhhhhh

SEQ    GRQLAERNGRPRRHRGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGA
SEG    ............................xxxxxxxxxxxxxxxxxxxxx..........
PRD    hhhhhhcccccccccccccccccccccccccccccccccccccccchhhhhhhhhccccc

SEQ    AEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAAVREERLHRPVQFCLLHRLLWLTW
SEG    ........................xxxxxxxxxxxxx.............xxxxxxxxxx
PRD    cccccccccccccccccccccccccccccccccccchhhhhhhhccccchhhhhhhhhhhc

SEQ    LPHPQAGGGGHQG
SEG    xxxxxxxxxxxxx
PRD    ccccccccccccc
```

(No Prosite data available for DKFZphtes3_18f3.2)

(No Pfam data available for DKFZphtes3_18f3.2)

Pedant information for DKFZphtes3_18f3, frame 3

Report for DKFZphtes3_18f3.3

```
[LENGTH]        248
[MW]            27162.56
[pI]            9.92
[PROSITE]       LEUCINE_ZIPPER 2
[KW]            TRANSMEMBRANE  1
[KW]            LOW_COMPLEXITY     30.65 %
[KW]            COILED_COIL        12.10 %

SEQ    MGMERPAAREPHGPDALRRFQGLLLDRRGRLHRQVLRLREVARRLERLRRRSLVANVAGS
SEG    ................xxxxxxxxxxxxxxxxxx.xxxxxxxxxxxxxxxxxxxx..xxx
PRD    cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccc
COILS  ............................................................
MEM    ............................................................

SEQ    SLSATGALAAIVGLSLSPVTLGTSLLVSAVGLGVATAGGAVTITSDLSLIFCNSRELRRV
SEG    xxxxxxxxxx..........xxxxxxxxxxxxxxxxxxxxxxxxx...............
PRD    cchhhhhhhhhhhhcccccccccccccccccceeeecccceeeeeeeceeeeeecchhhhhh
COILS  ............................................................
MEM    .......................MMMMMMMMMMMMMMMMMM...................

SEQ    QEIAATCQDQMREILSCLEFFCRWQGCGDRQLLQCGRNASIALYNSVYFIVFFGSRGFLI
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhccccchhhhhcceeeeeeecccccc
COILS  ............................................................
MEM    ............................................................

SEQ    PRRAEGDTKVSQAVLKAKIQKLAESLESCTGALDELSEQLESRVQLCTKSSRGHDLKISA
SEG    ............................................................
PRD    cccccccccchhhhhhhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhccccceeeehh
COILS  ............CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC..................
MEM    ............................................................

SEQ    DQRAGLFF
SEG    ........
PRD    hhhhhccc
COILS  ........
MEM    ........
```

Prosite for DKFZphtes3_18f3.3

PS00029	17->39	LEUCINE_ZIPPER	PDOC00029
PS00029	24->46	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_18f3.3)

DKFZphtes3_1817
----

group: cell structure and motility

DKFZphtes3_1817 encodes a novel 1050 amino acid protein with weak partial similarity to ankyrins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop) and an Ank repeat. Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus the novel protein seems to be involved in coupling of cyto skeleton and cell membrane.

The new protein can find application in modulation of cyto skeleton-membrane interactions.

similarity to ankyrins

Sequenced by MediGenomix

Locus: unknown

Insert length: 4501 bp
Poly A stretch at pos. 4423, no polyadenylation signal found

```
   1 GATCGCCGCG CGAGGGTGGT GGGCATCGAG GTCCCAGCAG CGGACGAGGG
  51 AGGTGCCGCC GTCGCCCAGG ATGGGCTGGG AATGAAGCGA TGTAGCCTTT
 101 TAAGAGATTT GCTCTGACCC ATCTGAAGTC CATATGGCTC TGTATGATGA
 151 AGACCTCCTG AAAAATCCTT TCTATCTGGC TCTGCAAAAG TGCCGCCCTG
 201 ACTTGTGCAG CAAAGTGGCC CAAATCCATG GCATTGTCTT AGTACCCTGC
 251 AAAGGAAGCC TGTCGAGCAG CATCCAGTCT ACTTGTCAGT TTGAGTCCTA
 301 CATTTTGATA CCTGTGGAAG AGCATTTTCA GACCTTAAAT GGAAAGGATG
 351 TCTTTATTCA AGGGAACAGG ATTAAATTAG GAGCTGGTTT TGCCTGTCTT
 401 CTCTCAGTGC CCATTCTCTT TGAAGAAACT TTCTACAATG AAAAAGAAGA
 451 GAGTTTCAGC ATCCTGTGTA TAGCCCATCC TTTGGAAAAG AGAGAGAGTT
 501 CAGAAGAGCC TTTGGCACCC TCAGATCCCT TTTCCCTGAA AACCATTGAA
 551 GATGTGAGAG AGTTCTTGGG AAGACACTCC GAGCGATTTG ACAGGAACAT
 601 CGCCTCTTTC CATCGAACAT TCCGAGAATG CGAGAGAAAG AGCCTCCGTC
 651 ACCACATAGA CTCAGCGAAT GCTCTCTACA CCAAATGCCT CCAGCAGCTT
 701 CTGAGGGACT CTCACCTGAA AATGCTCGCC AAGCAGGAGG CCCAGATGAA
 751 CCTGATGAAG CAGGCAGTGG AGATATACGT CCATCATGAA ATTTACAACC
 801 TGATCTTTAA ATACGTGGGG ACCATGGAGG CAAGTGAGGA TGCGGCCTTT
 851 AACAAAATCA CAAGAAGCCT TCAAGATCTT CAGCAGAAAG ATATTGGTGT
 901 GAAACCGGAG TTCAGCTTTA ACATACCTCG TGCCAAAAGA GAGCTGGCTC
 951 AGCTGAACAA ATGCACCTCC CCACAGCAGA AGCTTGTCTG CTTGCGAAAA
1001 GTGGTGCAGC TCATTACACA GTCTCCAAGC CAGAGAGTGA ACCTGGAGAC
1051 CATGTGTGCT GATGATCTGC TATCAGTCCT GTTATACTTG CTTGTGAAAA
1101 CGGAGATCCC TAATTGGATG GCAAATTTGA GTTACATCAA AAACTTCAGG
1151 TTTAGCAGCT TGGCAAAGGA TGAACTGGGA TACTGCCTGA CCTCATTCGA
1201 AGCTGCCATT GAATATATTC GGCAAGGAAG CCTCTCTGCT AAACCCCCTG
1251 AGTCTGAGGG ATTTGGAGAC AGGCTGTTCC TTAAGCAGAG AATGAGCTTA
1301 CTCTCTCAGA TGACTTCGTC TCCCACCGAC TGCCTGTTTA AGCACATTGC
1351 ATCAGGTAAC CAGAAAGAAG TGGAGAGACT TCTGAGCCAA GAGGACCATG
1401 ATAAAGATAC CGTCCAAAAG ATGTGTCACC CTCTCTGCTT CTGCGATGAC
1451 TGTGAGAAAC TCGTCTCTGG GAGGTTGAAT GATCCCTCAG TTGTCACTCC
1501 ATTCTCCAGA GACGACAGGG GGCACACCCC TCTCCATGTG GCTGCTGTCT
1551 GTGGGCAGGC ATCCCTCATC GACCTCCTGG TTTCCAAGGG CGCCATGGTA
1601 AATGCCACAG ACTACCATGG GGCCACTCCG CTCCACCTGG CCTGTCAGAA
1651 GGGCTACCAG AGCGTGACGC TGCTGCTGCT GCACTACAAG GCCAGCGCGG
1701 AAGTGCAGGA CAACAATGGG AATACGCCAC TCCACCTGGC CTGCACCTAC
1751 GGCCACGAGG ACTGTGTGAA GGCTCTGGTT TACTACGACG TGGAGTCGTG
1801 CAGACTTGAC ATTGGCAATG AGAAAGGAGA CACCCCTCTA CACATTGCTG
1851 CCCGCTGGGG CTACCAAGGC GTCATAGAGA CATTGCTGCA GAACGGAGCG
1901 TCCACCGAGA TCCAGAACAG ACTGAAGGAG ACGCCCCTCA AGTGTGCATT
1951 AAACTCAAAG ATTCTGTCTG TAATGGAAGC CTATCACCTG TCCTTCGAGA
2001 GGAGGCAGAA GTCGTCCGAG GCCCCTGTGC AGTCCCCGCA GCGCTCCGTG
2051 GACTCCATCA GCCAAGAGTC CTCCACTTCC AGCTTCTCCT CCATGTCAGC
2101 CGGCTCAAGG CAGGAGGAGA CCAAGAAGGA CTACAGAGAG GTAGAAAAAC
2151 TTTTGAGAGC AGTTGCTGAT GGAGATCTAG AAATGGTGCG TTACCTGTTG
2201 GAATGGACAG AGGAGGACCT GGAGGATGCG GAGGACACTG TCAGTGCAGC
2251 AGACCCCGAA TTCTGTCACC CGTTGTGCCA GTGCCCCAAG TGTGCCCCAG
2301 CTCAGAAGAG GCTGGCGAAG GTTCCTGCCA GTGGGCTTGG TGTGAACGTG
2351 ACCAGCCAGG ACGGCTCCTC CCCGCTGCAT GTCGCCGCCC TGCACGGCCG
2401 GGCGGACCTC ATCCGCCTCC TGCTGAAGCA CGGGGCCAAC GCAGGTGCCA
2451 GGAACGCAGA CCAAGCCGTC CCGCTCCACC TGGCCTGCCA GCAGGGCCAC
2501 TTTCAGGTGG TGAAGTGTCT GTTAGATTCG AATGCAAAAC CCAATAAGAA
2551 GGACCTCAGT GGAAACACGC CCCTCATTTA CGCCTGCTCC GGTGGCCATC
2601 ACGAGCTTGT GGCACTGCTG CTACAGCACG GGGCCTCCAT TAACGCTTCT
2651 AACAATAAGG GCAACACAGC GCTGCACGAG GCTGTGATTG AAAAGCACGT
```

```
2701 CTTCGTGGTA GAGCTGCTTC TGCTCCACGG AGCGTCAGTT CAGGTGCTGA
2751 ACAAGCGGCA GCGCACGGCT GTAGACTGTG CTGAACAGAA TTCAAAAATA
2801 ATGGAATTGC TTCAGGTGGT ACCAAGCTGT GTTGCTTCAT TAGATGATGT
2851 GGCTGAAACT GACCGCAAGG AGTATGTCAC TGTTAAGATC AGGAAAAAAT
2901 GGAACTCAAA ACTGTATGAT CTACCAGATG AGCCTTTTAC AAGACAGTTT
2951 TACTTTGTCC ACTCAGCTGG TCAGTTTAAG GGAAAGACTT CAAGGGAGAT
3001 TATGGCAAGA GATAGAAGTG TCCCTAATTT AACCGAAGGT TCTTTGCATG
3051 AGCCAGGGAG GCAAAGTGTC ACACTGAGAC AGAATAACCT GCCAGCTCAG
3101 AGTGGATCTC ATGCTGCTGA GAAAGGCAAC AGCGACTGGC CAGAGAGGCC
3151 TGGACTGACA CAGACTGGCC CTGGACACAG ACGGATGCTG CGGAGACACA
3201 CGGTAGAGGA TGCGGTCGTG TCCCAGGGCC CGGAGGCTGC TGGCCCCCTC
3251 TCCACTCCCC AAGAGGTTAG TGCTTCCCGG TCCTAACAGG AATGAGGAGT
3301 TGTTGAACCC ACTGCTAGGA AGCAAGGATG CAACAAGATG ATGCTGAGCG
3351 TGAACACATC TGAGAACTAA ATGTGCTTCC ATGAGACTGG CTTGAGAAGT
3401 CTTCAGCACC AAGTTCCTGA AAGCTTTTCT GTGGCAGGAA AGAATGCAAC
3451 AAAAAAGTTA ACCACCACCA TCTCTCTCCT CTTCAAAGCT AATGAATACA
3501 ATTGAAACAG ACAAAAATTC CAGTAGCATC CAGATCCTTA AGCCAGAGGT
3551 GCATGCTTCT TTTTAAGTAT GAGGGTTTGT TGGTCACAGT GGGAGAGGTT
3601 TCACCACCGC ATTCTGACCT CCTCCTCCCA AAAGGTGCTA AACCTCTCTG
3651 ACCTGTGTAC ATTCACAAAC CACAGCTAGA ATTCCTCCAC CTAGGATTAA
3701 GCTGGAGAGA AGTAAGTAAT TTAGGTTTCA TGGTACTGTA GAGGCCAGGC
3751 TGAAATGTCA TATCTGAAGG AAGAAAGCAG CAGCTGGACA ATGTTTCTTT
3801 GCAAAGCAAC ACTCGAACCA AAAGATGCCT CAATCCCATT TTGATATTCA
3851 TTTTAGTGAA AGGATGCATC AGACCTGTTC CACATCATGC ACATGGGAAA
3901 GGGTGGTTAT CATTTTCCTT CTAACAAGTA GGTACAGATA TTCGGTTACT
3951 ACACGTGCAC CTGTAGCAGT ATTTCTAGAA ACATCCCTTT TTGTTGAGAA
4001 CCTCCCTTGA ATGTCTGTCA CACTCACACC TGACGGGATG GTTACTGGAT
4051 TAGAGAGTAG ATTTGGCACA TCTTTTCTTA GTCTTTTGAT TCAAATTCAA
4101 AACTTAACAG CACAAACCAG GTCAGAGTTA CTTTCGGTTA GAATTTATTG
4151 CCATTTATTC CTTTTTATAA ATTTCTATAG ATTATACTGT TATTTTTATG
4201 TTATTGGCCT AGAGCTACAC GTATATGGGT TTGTCCTGAG TCCGTTTTCA
4251 AATGACCTTG TGATAGGGAA ATGGTTTTGT CCATGTTCTT GGAAATACTT
4301 GTGTATGTAC AGAAGGAAGG GAGGGATTAT TTTTCTACAA AGTAATTTAT
4351 GATTTCTAAT TTTCTAATGT GCCTTGGATA TGTGCCAAAT GATGGAAAAG
4401 AAACAGTAAA CTTTATGATT CTTAAAAAAA AAAAAAAAAA AAAAAAAAAA
4451 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAG
4501 G
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 3283 bp; peptide length: 1050
Category: similarity to known protein
Classification: Cell structure/motility
Prosite motifs: ATP_GTP_A (945-953)

```
  1 MALYDEDLLK NPFYLALQKC RPDLCSKVAQ IHGIVLVPCK GSLSSSIQST
 51 CQFESYILIP VEEHFQTLNG KDVFIQGNRI KLGAGFACLL SVPILFEETF
101 YNEKEESFSI LCIAHPLEKR ESSEEPLAPS DPFSLKTIED VREFLGRHSE
151 RFDRNIASFH RTFRECERKS LRHHIDSANA LYTKCLQQLL RDSHLKMLAK
201 QEAQMNLMKQ AVEIYVHHEI YNLIFKYVGT MEASEDAAFN KITRSLQDLQ
251 QKDIGVKPEF SFNIPRAKRE LAQLNKCTSP QQKLVCLRKV VQLITQSPSQ
301 RVNLETMCAD DLLSVLLYLL VKTEIPNWMA NLSYIKNFRF SSLAKDELGY
351 CLTSFEAAIE YIRQGSLSAK PPESEGFGDR LFLKQRMSLL SQMTSSPTDC
401 LFKHIASGNQ KEVERLLSQE DHDKDTVQKM CHPLCFCDDC EKLVSGRLND
451 PSVVTPFSRD DRGHTPLHVA AVCGQASLID LLVSKGAMVN ATDYHGATPL
501 HLACQKGYQS VTLLLLHYKA SAEVQDNNGN TPLHLACTYG HEDCVKALVY
551 YDVESCRLDI GNEKGDTPLH IAARWGYQGV IETLLQNGAS TEIQNRLKET
601 PLKCALNSKI LSVMEAYHLS FERRQKSSEA PVQSPQRSVD SISQESSTSS
651 FSSMSAGSRQ EETKKDYREV EKLLRAVADG DLEMVRYLLE WTEEDLEDAE
701 DTVSAADPEF CHPLCQCPKC APAQKRLAKV PASGLGVNVT SQDGSSPLHV
751 AALHGRADLI RLLLKHGANA GARNADQAVP LHLACQQGHF QVVKCLLDSN
801 AKPNKKDLSG NTPLIYACSG GHHELVALLL QHGASINASN NKGNTALHEA
851 VIEKHVFVVE LLLLHGASVQ VLNKRQRTAV DCAEQNSKIM ELLQVVPSCV
```

```
 901 ASLDDVAETD RKEYVTVKIR KKWNSKLYDL PDEPFTRQFY FVHSAGQFKG
 951 KTSREIMARD RSVPNLTEGS LHEPGRQSVT LRQNNLPAQS GSHAAEKGNS
1001 DWPERPGLTQ TGPGHRRMLR RHTVEDAVVS QGPEAAGPLS TPQEVSASRS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_1817, frame 2

TREMBL:HSU43965_1 gene: "ANK3"; product: "ankyrin G119"; Human ankyrin G119 (ANK3) mRNA, complete cds., N = 2, Score = 287, P = 3.7e-21

PIR:I49502 ankyrin - mouse, N = 3, Score = 365, P = 2.2e-27

TREMBL:HSANKY_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for ankyrin (variant 2.1), N = 2, Score = 380, P = 7.3e-31

SWISSPROT:ANK1_HUMAN ANKYRIN R (ANKYRINS 2.1 AND 2.2) (ERYTHROCYTE ANKYRIN)., N = 2, Score = 380, P = 8.2e-31

PIR:SJHUK ankyrin 1, erythrocyte splice form 1 - human, N = 2, Score = 380, P = 8.2e-31

```
>TREMBL:HSANKY_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for
     ankyrin (variant 2.1)
            Length = 1,719

  HSPs:

 Score = 380 (57.0 bits), Expect = 7.3e-31, Sum P(2) = 7.3e-31
 Identities = 139/447 (31%), Positives = 207/447 (46%)

Query:   462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521
             +G+T LH+AA+ GQ  ++  LV+ GA VNA     G TPL++A Q+ +  V    LL   A+
Sbjct:    77 KGNTALHIAALAGQDEVVRELVNYGANVNAQSQKGFTPLYMAAQENHLEVVKFLLENGAN 136

Query:   522 AEVQDNNGNTPLHLACTYGHEDCVKALVYYDVES-CRL---------------------- 558
               V    +G TPL +A    GHE+ V  L+ Y  +   RL
Sbjct:   137 QNVATEDGFTPLAVALQQGHENVVAHLINYGTKGKVRLPALHIAARNDDTRTAAVLLQND 196

Query:   559 ---DIGNEKGDTPLHIAARWGYQGVIETLLQNGASTEIQNRLKETPLKCALNSKILSVME 615
             D+ ++ G TPLHIAA +     V + LL  GAS    +   TPL  A  S+  +V+
Sbjct:   197 PNPDVLSKTGFTPLHIAAHYENLNVAQLLLNRGASVNFTPQNGITPLHIA--SRRGNVIM 254

Query:   616 AYHLSFERRQKSSEAPVQSPQRSVDSISQESSTS-SFSSMSAGSR-QEETKKDYREVEKL 673
              L  +R  +  E    +    ++        S  + G+  Q +TK      +
Sbjct:   255 V-RLLLDRGAQI-ETKTKDELTPLHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHM- 311

Query:   674 LRAVADGD-LEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAQKRLAKVPA 732
               A GD L+ VR LL++  E ++D  T+    P  H  C         R+AKV
Sbjct:   312 ---AAQGDHLDCVRLLLQYDAE-IDDI--TLDHLTP--LHVAAHC-----GHHRVAKVLL 358

Query:   733 S-GLGVNVTSQDGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVPLHLACQQGHFQ 791
              G  N  + +G +PLH+A       ++ LLLK GA+  A      PLH+A  GH
Sbjct:   359 DKGAKPNSRALNGFTPLHIACKKNHVRVMELLLKTGASIDAVTESGLTPLHVASFMGHLP 418

Query:   792 VVKCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALHEAV 851
             +VK LL   A PN  ++    TPL  A  GH E+    LLQ+ A +NA      T LH A
Sbjct:   419 IVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPLHCAA 478

Query:   852 IEKHVFVVELLLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVV 896
              H  +V+LLL + A+  +   T + A +   + +L ++
Sbjct:   479 RIGHTNMVKLLLENNANPNLATTAGHTPLHIAAREGHVETVLALL 523

 Score = 378 (56.7 bits), Expect = 1.2e-30, Sum P(2) = 1.2e-30
 Identities = 130/447 (29%), Positives = 195/447 (43%)

Query:   465 TPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEV 524
             TPLH AA  G   + ++L+ GA + A  +G +P+H+A Q +      LLL Y A +
Sbjct:   274 TPLHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHMAAQGDHLDCVRLLLQYDAEIDD 333

Query:   525 QDNNGNTPLHLACTYGHEDCVKALVYYDVE--------------SCR------------- 557
              + TPLH+A  GH   K L+  +           +C+
Sbjct:   334 ITLDHLTPLHVAAHCGHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMELLLKT 393

Query:   558 ---LDIGNEKGDTPLHIAARWGYQGVIETLLQNGASTEIQNRLKETPLKCALNSKILSVM 614
             +D  E G TPLH+A+ G+ +++ LLQ GAS + N  ETPL  A +    V
Sbjct:   394 GASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVA 453
```

```
Query:  615 EAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSSMSAGSRQEETKKDYREVEKLL 674
            + Y L  + +  +       Q+P        I   +        +A     T       L
Sbjct:  454 K-YLLQNKAKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGH----TPLH 508

Query:  675 RAVADGDLEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAQKRLAKVPASG 734
             A  +G +E V  LLE    ++    A  T     P   H    + K   A+  L +
Sbjct:  509 IAAREGHVETVLALLE---KEASQACMTKKGFTP--LHVAAKYGKVRVAELLLER----D 559

Query:  735 LGVNVTSQDGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVPLHLACQQGHFQVVK 794
              N    ++G +PLHVA  H   D+++LLL  G +  +     +    PLH+A +Q    +V +
Sbjct:  560 AHPNAAGKNGLTPLHVAVHHNNLDIVKLLLPRGGSPHSPAWNGYTPLHIAAKQNQVEVAR 619

Query:  795 CLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALHEAVIEK 854
             LL       N + + G TPL  A   GH E+VALLL   A+ N  N  G T LH     E
Sbjct:  620 SLLQYGGSANAESVQGVTPLHLAAQEGHAEMVALLLSKQANGNLGNKSGLTPLHLVAQEG 679

Query:  855 HVFVVELLLLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893
            HV V ++L+ HG  V     +   T +  A    N K+++ L
Sbjct:  680 HVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 720

 Score = 367 (55.1 bits), Expect = 1.8e-29, Sum P(2) = 1.8e-29
 Identities = 131/489 (26%), Positives = 210/489 (42%)

Query:  404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPL-CFCDDCEKLVSGRLNDPSVVTPFSRD 460
            HIAS  GN   V  LL +    +    PL C   +   +S  L D      ++
Sbjct:  244 HIASRRGNVIMVRLLLDRGAQIETKTKDELTPLHCAARNGHVRISEILLDHGAPIQ-AKT 302

Query:  461 DRGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKA 520
             G +P+H+AA          + LL+   A ++       TPLH+A   G+  V  +LL  A
Sbjct:  303 KNGLSPIHMAAQGDHLDCVRLLLQYDAEIDDITLDHLTPLHVAAHCGHHRVAKVLLDKGA 362

Query:  521 SAEVQDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPLHIAARWGYQGV 580
              +  NG TPLH+AC   H   ++ L++         +D  E G TPLH+A+  G+ +
Sbjct:  363 KPNSRALNGFTPLHIACKKNHVRVMELLLK---TGASIDAVTESGLTPLHVASFMGHLPI 419

Query:  581 IETLLQNGASTEIQNRLKETPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPVQSPQR 637
            ++ LLQ GAS  + N   ETPL  A   ++++   +                + K  + P+    R
Sbjct:  420 VKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPLHCAAR 479

Query:  638 ----SVDSISQESSTSSFSSMSAGSRQEETKKDYREVEKLLRAVADGDLEMVRYLLEWTE 693
             ++  +  E++ +  + +AG           VE +L  +         +T
Sbjct:  480 IGHTNMVKLLLENNANPNLATTAGHTPLHIAAREGHVETVLALLEKEASQACMTKKGFTP 539

Query:  694 EDLEDAEDTVSAAD---PEFCHPLCQ-----CP-KCAPAQKRLAKVPA---SGLGVNVTS 741
             +     V  A+      HP        P  A    L  V        G  +  +
Sbjct:  540 LHVAAKYGKVRVAELLLERDAHPNAAGKNGLTPLHVAVHHNNLDIVKLLLPRGGSPHSPA 599

Query:  742 QDGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCLLDSNA 801
             +G +PLH+AA   + ++  R LL++G +A A +    PLHLA Q+GH ++V  LL   A
Sbjct:  600 WNGYTPLHIAAKQNQVEVARSLLQYGGSANAESVQGVTPLHLAAQEGHAEMVALLLSKQA 659

Query:  802 KPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALHEAVIEKHVFVVEL 861
             N  + SG TPL     GH   +  +L++HG  ++A+  G T LH A     ++ +V+
Sbjct:  660 NGNLGNKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKF 719

Query:  862 LLLHGASVQVLNK 874
            LL H A V   K
Sbjct:  720 LLQHQADVNAKTK 732

 Score = 345 (51.8 bits), Expect = 4.2e-27, Sum P(2) = 4.2e-27
 Identities = 146/506 (28%), Positives = 233/506 (46%)

Query:  404 HIAS--GNQKEVERLLSQEDHDKDTVQK---MCHPLCFCDDCEKLVSGRLNDPSVVTPFS 458
            H+AS  G+ K V  LL +E    + T +K    H       +++V   +N  + V  +
Sbjct:   50 HLASKEGHVKMVVELLHKEIILETTTKKGNTALHIAALAGQ-DEVVRELVNYGANVN--A 106

Query:  459 RDDRGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHY 518
             +  +G TPL++AA     ++  L+  GA  N   G TPL +A Q+G+++V  L++Y
Sbjct:  107 QSQKGFTPLYMAAQENHLEVVKFLLENGANQNVATEDGFTPLAVALQQGHENVVAHLINY 166

Query:  519 KASAEVQDNNGNTP-LHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPLHIAARWGY 577
              +V+       P LH+A    ++D  A V    +   D+ ++ G TPLHIAA +
Sbjct:  167 GTKGKVR-----LPALHIAAR--NDDTRTAAVLLQNDP-NPDVLSKTGFTPLHIAAHYEN 218

Query:  578 QGVIETLLQNGASTEIQNRLKETPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPVQS 634
             V + LL  GAS     +   TPL  A   N ++ ++    E + K    P+
Sbjct:  219 LNVAQLLLNRGASVNFTPQNGITPLHIASRRGNVIMVRLLLDRGAQIETKTKDELTPLHC 278

Query:  635 PQRSVDSISQESSTSSFSSMSAGSRQEETKKDYREVEKLLRAVADGD-LEMVRYLLEWTE 693
             R+       E     + + A   +TK       +     A GD L+ VR LL++
Sbjct:  279 AARNGHVRISEILLDHGAPIQA-----KTKNGLSPIHM----AAQGDHLDCVRLLLQYDA 329
```

```
Query:   694 EDLEDAE-DTVSAAD-PEFC--HPLCQC-------PK-------------CAPAQKRLAK 729
             E ++D   D ++         C  H + +           P             C     R+ +
Sbjct:   330 E-IDDITLDHLTPLHVAAHCGHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVME 388

Query:   730 VPA-SGLGVNVTSQDGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVPLHLACQQG 788
             +    +G  ++   ++ G +PLHVA+  G     +++ LL+ GA+     N       PLH+A + G
Sbjct:   389 LLLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVKVETPLHMAARAG 448

Query:   789 HFQVVKCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALH 848
             H +V K LL + AK N K       TPL  A    GH  +V LLL++ A+ N +    G+T LH
Sbjct:   449 HTEVAKYLLQNKAKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTPLH 508

Query:   849 EAVIEKHVFVVELLLLHGASVQVLNKRQRTAVDCAEQNSKIM--ELL 893
               A  E HV  V  LL   AS   + K+  T +  A +  K+   ELL
Sbjct:   509 IAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAKYGKVRVAELL 555

 Score = 243 (36.5 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14
 Identities = 64/199 (32%), Positives = 97/199 (48%)

Query:   404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFCDDCEKLVSGRLNDPSVVTPFSRDD 461
             H+A++  G  +  E LL ++ H      +     PL          L    +L  P   +P S
Sbjct:   541 HVAAKYGKVRVAELLLERDAHPNAAGKNGLTPLHVAVHHNNLDIVKLLLPRGGSPHSPAW 600

Query:   462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521
              G+TPLH+AA    Q  +   L+  G    NA     G TPLHLA Q+G+  +  LLL   +A+
Sbjct:   601 NGYTPLHIAAKQNQVEVARSLLQYGGSANAESVQGVTPLHLAAQEGHAEMVALLLSKQAN 660

Query:   522 AEVQDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPLHIAARWGYQGVI 581
               + + +G TPLHL     GH       L+ + V     +D      G TPLH+A++ +G    ++
Sbjct:   661 GNLGNKSGLTPLHLVAQEGHVPVADVLIKHGV---MVDATTRMGYTPLHVASHYGNIKLV 717

Query:   582 ETLLQNGASTEIQNRLKETPL 602
             + LLQ+ A    + +L  +PL
Sbjct:   718 KFLLQHQADVNAKTKLGYSPL 738

 Score = 242 (36.3 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29
 Identities = 63/176 (35%), Positives = 92/176 (52%)

Query:   734 GLGVNVTSQDGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVPLHLACQQGHFQVV 793
             G  VN T Q+G +PLH+A++  G    ++RLLL  GA     +  D+  PLH A + GH ++
Sbjct:   229 GASVNFTPQNGITPLHIASRRGNVIMVRLLLDRGAQIETKTKDELTPLHCAARNGHVRIS 288

Query:   794 KCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALHEAVIE 853
             + LLD  A     K  +G +P+  A  G H + V LLLQ+ A I+         T LH A
Sbjct:   289 EILLDHGAPIQAKTKNGLSPIHMAAQGDHLDCVRLLLQYDAEIDDITLDHLTPLHVAAHC 348

Query:   854 KHVFVVELLLLHGA--SVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAET 909
                H  V ++LL  GA  + + LN      + C + + ++MELL         AS+D V E+
Sbjct:   349 GHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMELLLKTG---ASIDAVTES 403

 Score = 242 (36.3 bits), Expect = 3.3e-14, Sum P(2) = 3.3e-14
 Identities = 80/284 (28%), Positives = 129/284 (45%)

Query:   404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFCDDCEKLVSGRLNDPSVVTPFSRDD 461
             HIA+  G+ + V  LL +E        +K   PL         K+    L      P +
Sbjct:   508 HIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAAGK 567

Query:   462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521
              G TPLHVA        ++ LL+ +G    ++  ++G TPLH+A ++    V    LL Y  S
Sbjct:   568 NGLTPLHVAVHHNNLDIVKLLLPRGGSPHSPAWNGYTPLHIAAKQNQVEVARSLLQYGGS 627

Query:   522 AEVQDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPLHIAARWGYQGVI 581
             A   +    G TPLHLA    GH + V  L+        ++GN+ G TPLH+ A+ G+  V
Sbjct:   628 ANAESVQGVTPLHLAAQEGHAEMVALLLSKQANG---NLGNKSGLTPLHLVAQEGHVPVA 684

Query:   582 ETLLQNGASTEIQNRLKETPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPV-QSPQR 637
             + L+++G    +    R+  TPL  A    N K++  +  +  +  K    +P++ Q+ Q+
Sbjct:   685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLLQHQADVNAKTKLGYSPLHQAAQQ 744

Query:   638 S-VDSISQ--ESSTSSFSSMSAGSRQEETKK--DYREVEKLLRAVAD 679
               D  ++    ++    S    S  G+    K    Y  V  +L+  V D
Sbjct:   745 GHTDIVTLLLKNGASPNEVSSDGTTPLAIAKRLGYISVTDVLKVVTD 791

 Score = 235 (35.3 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34
 Identities = 58/165 (35%), Positives = 83/165 (50%)

Query:   734 GLGVNVTSQDGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVPLHLACQQGHFQVV 793
             G    N   S    G +PLH+AA  G A+++ LLL   AN     N     PLHL  Q+GH  V
Sbjct:   625 GGSANAESVQGVTPLHLAAQEGHAEMVALLLSKQANGNLGNKSGLTPLHLVAQEGHVPVA 684
```

```
Query:   794 KCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALHEAVIE 853
             L+       +       G TPL  A    G+ +LV  LLQH A +NA    G + LH+A  +
Sbjct:   685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLLQHQADVNAKTKLGYSPLHQAAQQ 744

Query:   854 KHVFVVELLLLHGASVQVLNKRQRTAVDCAEQNS--KIMELLQVV 896
              H  +V LLL +GAS    ++      T +  A++       + ++L+VV
Sbjct:   745 GHTDIVTLLLKNGASPNEVSSDGTTPLAIAKRLGYISVTDVLKVV 789

 Score = 233 (35.0 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34
 Identities = 67/202 (33%), Positives = 100/202 (49%)

Query:   404 HIAS-GNQKEVERLLSQEDHDKDTVQKMCH--PLCFCDDC-EKLVSGRLNDPSVVTPFSR 459
             H+A++ G+  +  RLL Q D + D +  + H  PL      C     V+  L D     P SR
Sbjct:   310 HMAAQGDHLDCVRLLLQYDAEIDDIT-LDHLTPLHVAAHCGHHRVAKVLLDKGA-KPNSR 367

Query:   460 DDRGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519
               G TPLH+A        +++LL+  GA ++A    G TPLH+A   G+  +    LL
Sbjct:   368 ALNGFTPLHIACKKNHVRVMELLLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRG 427

Query:   520 ASAEVQDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPLHIAARWGYQG 579
             AS  V +      TPLH+A    GH +  K L+          +++   +    TPLH AAR G+
Sbjct:   428 ASPNVSNVKVETPLHMAARAGHTEVAKYLLQ---NKAKVNAKAKDDQTPLHCAARIGHTN 484

Query:   580 VIETLLQNGASTEIQNRLKETPLKCA 605
             +++ LL+N A+  +        TPL  A
Sbjct:   485 MVKLLLENNANPNLATTAGHTPLHIA 510

 Score = 226 (33.9 bits), Expect = 7.0e-33, Sum P(2) = 7.0e-33
 Identities = 53/153 (34%), Positives = 83/153 (54%)

Query:   743 DGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCLLDSNAK 802
             +G +PLH+AA    + ++ R LL++G +A A +      PLHLA Q+GH ++V  LL    A
Sbjct:   601 NGYTPLHIAAKQNQVEVARSLLQYGGSANAESVQGVTPLHLAAQEGHAEMVALLLSKQAN 660

Query:   803 PNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALHEAVIEKHVFVVELL 862
             N  + SG TPL         GH   +  +L++HG  ++A+   G T LH A     ++ +V+ L
Sbjct:   661 GNLGNKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 720

Query:   863 LLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893
             L H A V    K   + +  A Q  ++ I+ LL
Sbjct:   721 LQHQADVNAKTKLGYSPLHQAAQQGHTDIVTLL 753

 Score = 198 (29.7 bits), Expect = 2.5e-11, Sum P(2) = 2.5e-11
 Identities = 51/157 (32%), Positives = 82/157 (52%)

Query:   737 VNVTSQDGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCL 796
             +  T++ G++ LH+AAL G+ +++R L+ +GAN  A++      PL++A Q+ H +VVK L
Sbjct:    71 LETTTKKGNTALHIAALAGQDEVVRELVNYGANVNAQSQKGFTPLYMAAQENHLEVVKFL 130

Query:   797 LDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALHEAVIEKHV 856
             L++ A  N      G TPL  A    GH  +VA L+ +G              ALH A
Sbjct:   131 LENGANQNVATEDGFTPLAVALQQGHENVVAHLINYGTK----GKVRLPALHIAARNDDT 186

Query:   857 FVVELLLLHGASVQVLNKRQRTAVDCAE--QNSKIMELL 893
                +LL +   +  VL+K     T +  A    +N  + +LL
Sbjct:   187 RTAAVLLQNDPNPDVLSKTGFTPLHIAAHYENLNVAQLL 225

 Score = 186 (27.9 bits), Expect = 6.6e-29, Sum P(2) = 6.6e-29
 Identities = 55/143 (38%), Positives = 68/143 (47%)

Query:   463 GHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASA 522
             GHTPLH+AA   G     +  L+ K A          G TPLH+A + G   V LLL   A
Sbjct:   503 GHTPLHIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHP 562

Query:   523 EVQDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPLHIAARWGYQGVIE 582
                  NG TPLH+A  + + D VK L+      S        N  G TPLHIAA+        V
Sbjct:   563 NAAGKNGLTPLHVAVHHNNLDIVKLLLPRG-GSPHSPAWN--GYTPLHIAAKQNQVEVAR 619

Query:   583 TLLQNGASTEIQNRLKETPLKCA 605
             +LLQ G S    ++      TPL  A
Sbjct:   620 SLLQYGGSANAESVQGVTPLHLA 642

 Score = 182 (27.3 bits), Expect = 2.9e-28, Sum P(2) = 2.9e-28
 Identities = 54/185 (29%), Positives = 89/185 (48%)

Query:   738 NVTSQDGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCLL 797
             N+ ++ G +PLH+ A  G    +  +L+KHG    A          PLH+A    G+ ++VK LL
Sbjct:   662 NLGNKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLL 721

Query:   798 DSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALHEAVIEKHVF 857
                 A  N K     G +PL  A     GH ++V LLL++GAS N    ++ G T L  A     ++
```

```
Sbjct:   722 QHQADVNAKTKLGYSPLHQAAQQGHTDIVTLLLKNGASPNEVSSDGTTPLAIAKRLGYIS 781

Query:   858 VVELLLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAETDRKEYVTV 917
             V ++L +      V ++      V     + S           P  V  + DV+E + +E ++
Sbjct:   782 VTDVLKV------VTDETSFVLVSDKHRMS--------FPETVDEILDVSEDEGEELISF 827

Query:   918 KIRKK 922
             K  ++
Sbjct:   828 KAERR 832

 Score = 180 (27.0 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29
 Identities = 41/121 (33%), Positives = 67/121 (55%)

Query:   486 GAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEVQDNNGNTPLHLACTYGHEDCV 545
             G  +N  + +G    LHLA ++G+  + + LLH +    E      GNT LH+A    G ++ V
Sbjct:    35 GVDINTCNQNGLNGLHLASKEGHVKMVVELLHKEIILETTKKGNTALHIAALAGQDEVV 94

Query:   546 KALVYYDVESCRLDIGNEKGDTPLHIAARWGYQGVIETLLQNGASTEIQNRLKETPLKCA 605
             + LV Y         ++   ++KG TPL++AA+  +  V++ LL+NGA+  +        TPL  A
Sbjct:    95 RELVNY---GANVNAQSQKGFTPLYMAAQENHLEVVKFLLENGANQNVATEDGFTPLAVA 151

Query:   606 L 606
             L
Sbjct:   152 L 152

 Score = 166 (24.9 bits), Expect = 3.4e-06, Sum P(2) = 3.4e-06
 Identities = 89/318 (27%), Positives = 140/318 (44%)

Query:   448 LNDPSVVTPFSRDDRGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKG 507
             L + + V    ++DD+  TPLH AA  G  +++ LL+   A  N     G TPLH+A ++G
Sbjct:   457 LQNKAKVNAKAKDDQ--TPLHCAARIGHTNMVKLLLENNANPNLATTAGHTPLHIAAREG 514

Query:   508 YQSVTLLLLHYKASAEVQDNNGNTPLHLACTYGHEDCVKALVYYD--------------- 552
             +    L LL   +AS          G TPLH+A  YG       + L+  D
Sbjct:   515 HVETVLALLEKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAAGKNGLTPLH 574

Query:   553 --VESCRLDI--------GNE-----KGDTPLHIAARWGYQGVIETLLQNGASTEIQNRL 597
               V    LDI          G+       G TPLHIAA+      V  +LLQ G S    ++
Sbjct:   575 VAVHHNNLDIVKLLLPRGGSPHSPAWNGYTPLHIAAKQNQVEVARSLLQYGGSANAESVQ 634

Query:   598 KETPLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSSM-SA 656
               TPL  A        M A  LS    +Q +      +S     +  ++QE         +
Sbjct:   635 GVTPLHLAAQEGHAE-MVALLLS---KQANGNLGNKSGLTPLHLVAQEGHVPVADVLIKH 690

Query:   657 GSRQEETKKDYREVEKLLRAVADGDLEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQ 716
             G   + T +           L  A   G++++V++LL+  + D+ +A++  +  +      PL Q
Sbjct:   691 GVMVDATTR--MGYTPLHVASHYGNIKLVKFLLQH-QADV-NAKTKLGYS------PLHQ 740

Query:   717 CPKCAPAQKRLAKVPASGLGVNVTSQDGSSPLHVA 751
             +       +  +  +G   N   S DG++PL +A
Sbjct:   741 AAQQGHTDI-VTLLLKNGASPNEVSSDGTTPLAIA 774

 Score = 162 (24.3 bits), Expect = 1.8e-07, Sum P(2) = 1.8e-07
 Identities = 48/149 (32%), Positives = 71/149 (47%)

Query:   737 VNVTSQDGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVPLHLACQQGHFQVVKCL 796
             V        D ++        AA  G  D       L++G +      N +       LHLA ++GH ++V  L
Sbjct:     5 VGFREADAATSFLRAARSGNLDKALDHLRNGVDINTCNQNGLNGLHLASKEGHVKMVVEL 64

Query:   797 LDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASNNKGNTALHEAVIEKHV 856
             L                GNT L  A    G  E+V  L+ +GA++NA + KG T L+ A  E H+
Sbjct:    65 LHKEIILETTKKGNTALHIAALAGQDEVVRELVNYGANVNAQSQKGFTPLYMAAQENHL 124

Query:   857 FVVELLLLHGASVQVLNKRQRTAVDCAEQ 885
             VV+ LL +GA+  V  +    T + A Q
Sbjct:   125 EVVKFLLENGANQNVATEDGFTPLAVALQ 153

 Score = 158 (23.7 bits), Expect = 5.7e-26, Sum P(2) = 5.7e-26
 Identities = 38/135 (28%), Positives = 65/135 (48%)

Query:   460 DDRGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519
             + G   LH+A++  G    ++  L+ K  ++  T    G T LH+A    G    V     L+++Y
Sbjct:    42 NQNGLNGLHLASKEGHVKMVVELLHKEIILETTTKKGNTALHIAALAGQDEVVRELVNYG 101

Query:   520 ASAEVQDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPLHIAARWGYQG 579
             A+     Q     G TPL++A        H + VK L+              ++  E G TPL +A + G++
Sbjct:   102 ANVNAQSQKGFTPLYMAAQENHLEVVKFLLE---NGANQNVATEDGFTPLAVALQQGHEN 158

Query:   580 VIETLLQNGASTEIQ 594
             V+ L+  G +++
Sbjct:   159 VVAHLINYGTKGKVR 173
```

```
 Score = 115 (17.3 bits), Expect = 1.8e-21, Sum P(2) = 1.8e-21
 Identities = 37/119 (31%), Positives = 58/119 (48%)

Query:   497 ATPLHLACQKGYQSVTLLLLHYKASAEVQ--DNNGNTPLHLACTYGHEDCVKALVYYDVE 554
             AT    A + G  ++    L H +    ++    + NG    LHLA    GH    V   L++ ++
Sbjct:    13 ATSFLRAARSG--NLDKALDHLRNGVDINTCNQNGLNGLHLASKEGHVKMVVELLHKEII 70

Query:   555 SCRLDIGNEKGDTPLHIAARWGYQGVIETLLQNGASTEIQNRLKETPLKCALNSKILSVM 614
                L+   +KG+T LHIAA  G    V+  L+  GA+    Q++    TPL  A      L V+
Sbjct:    71 ---LETTTKKGNTALHIAALAGQDEVVRELVNYGANVNAQSQKGFTPLYMAAQENHLEVV 127

Query:   615 E 615
             +
Sbjct:   128 K 128

 Score = 106 (15.9 bits), Expect = 1.8e-01, Sum P(2) = 1.6e-01
 Identities = 34/121 (28%), Positives = 54/121 (44%)

Query:   769 NAGARNADQAVPLHLACQQGHFQVVKCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVAL 828
             + G R AD A      A + G+             L +      N  + +G    L  A     GH ++V
Sbjct:     4 SVGFREADAATSFLRAARSGNLDKALDHLRNGVDINTCNQNGLNGLHLASKEGHVKMVVE 63

Query:   829 LLQHGASINASNNKGNTALHEAVIEKHVFVVELLLLHGASVQVLNKRQRTAVDCAEQNSK 888
             LL       +   +   KGNTALH  A  +        VV   L+  +GA+V      +++   T +   A Q +
Sbjct:    64 LLHKEIILETTTKKGNTALHIAALAGQDEVVRELVNYGANVNAQSQKGFTPLYMAAQENH 123

Query:   889 I 889
             +
Sbjct:   124 L 124

 Score = 40 (6.0 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14
 Identities = 11/56 (19%), Positives = 23/56 (41%)

Query:   622 ERRQKSSEAPVQSPQRSVDSISQESSTSSFSSMSAGSRQEETKKDYREVEKLLRAV 677
             +RRQ+   E  VQ  + +   + Q +    +              Q   ++       +K++R V
Sbjct:  1614 DRRQQGQEEQVQEAKNTFTQVVQGNEFQNIPGEQVTEEQFTDEQGNIVTKKIIRKV 1669

 Score = 38 (5.7 bits), Expect = 2.6e-14, Sum P(2) = 2.6e-14
 Identities = 6/12 (50%), Positives = 10/12 (83%)

Query:   806 KDLSGNTPLIYA 817
             +D++G T L+YA
Sbjct:  1186 EDITGTTKLVYA 1197
```

Pedant information for DKFZphtes3_1817, frame 2

Report for DKFZphtes3_1817.2

```
[LENGTH]        1050
[MW]            117013.72
[pI]            6.47
[HOMOL]         TREMBL:DMANKY_1 product: "ankyrin"; Drosophila melanogaster ankyrin mRNA,
complete cds. 2e-45
[FUNCAT]        08.19 cellular import  [S. cerevisiae, YOR034c] 5e-13
[FUNCAT]        10.05.99 other pheromone response activities       [S. cerevisiae, YDR264c]
3e-12
[FUNCAT]        03.07 pheromone response, mating-type determination, sex-specific proteins
        [S. cerevisiae, YDR264c] 3e-12
[FUNCAT]        99 unclassified proteins      [S. cerevisiae, YIL112w] 2e-11
[FUNCAT]        06.13.01 cytoplasmic degradation      [S. cerevisiae, YGR232w] 8e-10
[FUNCAT]        30.10 nuclear organization      [S. cerevisiae, YIR033w] 2e-08
[FUNCAT]        04.05.01.07 chromatin modification    [S. cerevisiae, YIR033w] 2e-08
[FUNCAT]        01.04.04 regulation of phosphate utilization       [S. cerevisiae, YGR233c]
3e-08
[FUNCAT]        08.13 vacuolar transport      [S. cerevisiae, YML097c] 5e-05
[FUNCAT]        06.04 protein targeting, sorting and translocation  [S. cerevisiae, YML097c]
5e-05
[FUNCAT]        30.03 organization of cytoplasm       [S. cerevisiae, YML097c] 5e-05
[FUNCAT]        08.07 vesicular transport (golgi network, etc.)     [S. cerevisiae, YML097c]
5e-05
[FUNCAT]        03.22 cell cycle control and mitosis [S. cerevisiae, YER111c] 3e-04
[FUNCAT]        04.05.01.04 transcriptional control  [S. cerevisiae, YER111c] 3e-04
[BLOCKS]        BL00901A Cysteine synthase/cystathionine beta-synthase P-phosphate att
[SCOP]          d1awcb_ 1.91.3.1.2 GA binding protein (GABP) alpha GA bindini 4e-12
[EC]            3.1.3.53 Myosin-light-chain-phosphatase 1e-12
[PIRKW]         phosphotransferase 1e-19
[PIRKW]         nucleus 1e-13
```

```
[PIRKW]        potassium channel 5e-15
[PIRKW]        early protein 2e-13
[PIRKW]        tumor suppressor 1e-09
[PIRKW]        duplication 1e-14
[PIRKW]        tandem repeat 1e-19
[PIRKW]        heterodimer 1e-14
[PIRKW]        potassium transport 5e-15
[PIRKW]        cell cycle control 1e-10
[PIRKW]        serine/threonine-specific protein kinase 1e-19
[PIRKW]        transmembrane protein 5e-15
[PIRKW]        transport protein 5e-15
[PIRKW]        DNA binding 2e-11
[PIRKW]        oncogene 1e-08
[PIRKW]        ATP 1e-19
[PIRKW]        protein kinase inhibitor 1e-09
[PIRKW]        voltage-gated ion channel 5e-15
[PIRKW]        phosphoprotein 4e-38
[PIRKW]        apoptosis 1e-19
[PIRKW]        liver 4e-09
[PIRKW]        integrin binding 3e-16
[PIRKW]        differentiation 2e-12
[PIRKW]        transforming protein 1e-08
[PIRKW]        alternative splicing 1e-40
[PIRKW]        coiled coil 1e-14
[PIRKW]        peripheral membrane protein 2e-38
[PIRKW]        transcription factor 4e-16
[PIRKW]        transcription regulation 2e-16
[PIRKW]        nucleotide binding 5e-15
[PIRKW]        phosphoric monoester hydrolase 1e-12
[PIRKW]        cytoskeleton 8e-39
[PIRKW]        calmodulin binding 1e-19
[PIRKW]        smooth muscle 1e-12
[SUPFAM]       ankyrin 1e-40
[SUPFAM]       death-associated protein kinase 1e-19
[SUPFAM]       ankyrin repeat homology 1e-40
[SUPFAM]       protein kinase homology 1e-19
[SUPFAM]       vaccinia virus 27.4K HindIII-C protein homology 3e-07
[SUPFAM]       int-3 transforming protein 1e-08
[SUPFAM]       unassigned ankyrin repeat proteins 2e-38
[SUPFAM]       notch protein 2e-12
[SUPFAM]       fowlpox virus BamHI-ORF7 protein 2e-13
[SUPFAM]       rel homology 2e-11
[SUPFAM]       EGF homology 2e-12
[PROSITE]      ATP_GTP_A      1
[PFAM]         Ank repeat
[KW]           Irregular
[KW]           3D
[KW]           LOW_COMPLEXITY     3.05 %

SEQ    MALYDEDLLKNPFYLALQKCRPDLCSKVAQIHGIVLVPCKGSLSSSIQSTCQFESYILIP
SEG    ............................................................
1awcB  ............................................................

SEQ    VEEHFQTLNGKDVFIQGNRIKLGAGFACLLSVPILFEETFYNEKEESFSILCIAHPLEKR
SEG    ............................................................
1awcB  ............................................................

SEQ    ESSEEPLAPSDPFSLKTIEDVREFLGRHSERFDRNIASFHRTFRECERKSLRHHIDSANA
SEG    ............................................................
1awcB  ............................................................

SEQ    LYTKCLQQLLRDSHLKMLAKQEAQMNLMKQAVEIYVHHEIYNLIFKYVGTMEASEDAAFN
SEG    ............................................................
1awcB  ............................................................

SEQ    KITRSLQDLQQKDIGVKPEFSFNIPRAKRELAQLNKCTSPQQKLVCLRKVVQLITQSPSQ
SEG    ............................................................
1awcB  ............................................................

SEQ    RVNLETMCADDLLSVLLYLLVKTEIPNWMANLSYIKNFRFSSLAKDELGYCLTSFEAAIE
SEG    ...........xxxxxxxxxx.......................................
1awcB  ............................................................

SEQ    YIRQGSLSAKPPESEGFGDRLFLKQRMSLLSQMTSSPTDCLFKHIASGNQKEVERLLSQE
SEG    ............................................................
1awcB  ............................................................

SEQ    DHDKDTVQKMCHPLCFCDDCEKLVSGRLNDPSVVTPFSRDDRGHTPLHVAAVCGQASLID
SEG    ............................................................
1awcB  ............................................................
```

```
SEQ     LLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEVQDNNGNTPLHLACTYG
SEG     ............................................................
lawcB   ............................................................

SEQ     HEDCVKALVYYDVESCRLDIGNEKGDTPLHIAARWGYQGVIETLLQNGASTEIQNRLKET
SEG     ............................................................
lawcB   ............................................................

SEQ     PLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSSMSAGSRQ
SEG     .....................................xxxxxxxxxxxxxxxxxxxxxx.
lawcB   ............................................................

SEQ     EETKKDYREVEKLLRAVADGDLEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKC
SEG     ............................................................
lawcB   ............................................................

SEQ     APAQKRLAKVPASGLGVNVTSQDGSSPLHVAALHGRADLIRLLLKHGANAGARNADQAVP
SEG     ............................................................
lawcB   .....................CHHHHHHHHHHHCCHHHHHHHHHHHCCCC-CCTTTTCCH

SEQ     LHLACQQGHFQVVKCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQHGASINASN
SEG     ............................................................
lawcB   HHHHHHHCCHHHHHHHHHCCCTTTTCTTTTCCHHHHHHHHTTHHHHHHHHHCCCTTTTEE

SEQ     NKGNTALHEAVIEKHVFVVELLLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCV
SEG     ............................................................
lawcB   TTTEEHHHHHHHHCCHHHHHHHHHHHCCTTTTCBTTTBCHHHHHHHHHCCHHHHHC......

SEQ     ASLDDVAETDRKEYVTVKIRKKWNSKLYDLPDEPFTRQFYFVHSAGQFKGKTSREIMARD
SEG     ............................................................
lawcB   ............................................................

SEQ     RSVPNLTEGSLHEPGRQSVTLRQNNLPAQSGSHAAEKGNSDWPERPGLTQTGPGHRRMLR
SEG     ............................................................
lawcB   ............................................................

SEQ     RHTVEDAVVSQGPEAAGPLSTPQEVSASRS
SEG     ..............................
lawcB   ..............................
```

Prosite for DKFZphtes3_1817.2

```
PS00017     945->953    ATP_GTP_A          PDOC00017
```

Pfam for DKFZphtes3_1817.2

```
HMM_NAME        Ank repeat

HMM                   *GyTPLHIAARyNNvEMVrlLLQHGADIN*
                       G+TPLH+AA  ++ ++++LL+++GA +N
Query            463   GHTPLHVAAVCGQASLIDLLVSKGAMVN     490

32.12  (bits) f:  496 t:  523 Target: dkfzphtes3_1817.2 similarity to ankyrins
  Alignment to HMM consensus:
Query                 *GyTPLHIAARyNNvEMVrlLLQHGADIN*
                       G TPLH+A++ +  ++  LLL + A+
  dkfzphtes3     496   GATPLHLACQKGYQSVTLLLLHYKASAE     523

Query           f:  529 t:  556 Target: dkfzphtes3_1817.2 similarity to ankyrins
  Alignment to HMM consensus:
HMM                   *GyTPLHIAARyNNvEMVrlLLQHGADIN*
                       G+TPLH+A+ Y+++++V+ L+ +
Query            529   GNTPLHLACTYGHEDCVKALVYYDVESC     556

42.65  (bits) f:  565 t:  592 Target: dkfzphtes3_1817.2 similarity to ankyrins
  Alignment to HMM consensus:
Query                 *GyTPLHIAARyNNvEMVrlLLQHGADIN*
                       G+TPLHIAAR +   +++ LLQ+GA+
  dkfzphtes3     565   GDTPLHIAARWGYQGVIETLLQNGASTE     592

Query           f:  744 t:  771 Target: dkfzphtes3_1817.2 similarity to ankyrins
  Alignment to HMM consensus:
HMM                   *GyTPLHIAARyNNvEMVrlLLQHGADIN*
                       G +PLH+AA +++ +++RLLL+HGA+
Query            744   GSSPLHVAALHGRADLIRLLLKHGANAG     771
```

```
36.38 (bits) f: 777 t: 804 Target: dkfzphtes3_1817.2 similarity to ankyrins
  Alignment to HMM consensus:
Query                 *GyTPLHIAARyNNvEMVrlLLQHGADIN*
                          PLH+A+++++ ++V+ LL+ +A +N
  dkfzphtes3   777  QAVPLHLACQQGHFQVVKCLLDSNAKPN    804

Query          f:  810 t:  837 Target: dkfzphtes3_1817.2 similarity to ankyrins
  Alignment to HMM consensus:
HMM                   *GyTPLHIAARyNNvEMVrlLLQHGADIN*
                      G+TPL++A+   ++ E+V LLLQHGA+IN
Query          810  GNTPLIYACSGGHHELVALLLQHGASIN    837

44.62  (bits) f:  843 t:  870 Target: dkfzphtes3_1817.2 similarity to ankyrins
  Alignment to HMM consensus:
Query                 *GyTPLHIAARyNNvEMVrlLLQHGADIN*
                      G+T+LH A+++ +V +V+LLL HGA++
  dkfzphtes3   843  GNTALHEAVIEKHVFVVELLLLHGASVQ    870
```

DKFZphtes3_19f19
----

group: testes derived

DKFZphtes3_19f19 encodes a novel 254 amino acid protein with weak similarity to S. cerevisiae protein YFL046w.

The protein contains a RGD cell attachment site.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YFL046w

localisation: 3 STS match perfect but HS1292427 matches to chromosome 4

Sequenced by MediGenomix

Locus: /map="405.0/.3 cR from top of Chr11 linkage group"

Insert length: 1395 bp
Poly A stretch at pos. 1367, no polyadenylation signal found

```
   1 GGGACCACGG TGGCGCCTGC GCTGGGAGGT GAGCTTGTGA CAGAGCGAAA
  51 ACTACAATTC CCAGCATTCC TGTGGTGCCA GAACTACCTT GCCCGAAAGC
 101 CTGTGCGAGA TTTACCCCGT CTTCCGCCTC CCTCCCACCG GAAAACTCTG
 151 AGGACATGAA TAGTCGCCAG GCTTGGCGGC TCTTTCTCTC CCAAGGCAGA
 201 GGAGATCGTT GGGTTTCAAG GCCCCGCGGG CATTTCTCGC CGGCCCTGCG
 251 GAGAGAGTTC TTCACTACCA CAACCAAGGA GGGATATGAT AGGCGGCCAG
 301 TGGATATAAC TCCTTTAGAA CAAAGGAAAT TAACTTTTGA TACCCATGCA
 351 TTGGTTCAGG ACTTGGAAAC TCATGGATTT GACAAAACAC AAGCAGAAAC
 401 AATTGTATCA GCGTTAACTG CTTTATCAAA TGTCAGCCTG GATACTATCT
 451 ATAAAGAGAT GGTCACTCAA GCTCAACAGG AAATAACAGT ACAACAGCTA
 501 ATGGCTCATT TGGATGCTAT CAGGAAAGAC ATGGTCATCC TAGAGAAAAG
 551 TGAATTTGCA AATCTGAGAG CAGAGAATGA GAAAATGAAA ATTGAATTAG
 601 ACCAAGTTAA GCAACAACTA ATGCATGAAA CCAGTCGAAT CAGAGCAGAT
 651 AATAAACTGG ATATCAACTT AGAAAGGAGC AGAGTAACAG ATATGTTTAC
 701 AGATCAAGAA AAGCAACTTA TGGAAACAAC TACAGAATTT ACAAAAAAGG
 751 ATACTCAAAC CAAAAGTATT ATTTCAGAGA CCAGTAATAA AATTGACGCT
 801 GAAATTGCTT CCTTAAAAAC ACTGATGGAA TCTAACAAAC TTGAGACAAT
 851 TCGTTATCTT GCAGCTTCGG TGTTTACTTG CCTGGCAATA GCATTGGGAT
 901 TTTATAGATT CTGGAAGTAG TATTAATGCT CATCCTGCTG TGGCTGTTGG
 951 CTTCTTAGAA CACCAAACCG GGAGAGATTT ACTTTGAACA TTGTCAGTTG
1001 CAGCAAAAAT TTACTACACA AGATTATTCG AAGTGTATAC GGACTAAAAG
1051 AGGAAGTGTT TTAGAATGAG AAGAGATACT GTGTCTTTAT TGTGTGTGTG
1101 TGAGTGCAGG TGTGTGTCTT TATTATATTG AAAAGCTGTC ACTCAGACCT
1151 GGTTTGAGAT AGAAGAGCAT TTTGTCCTTT TGATAGTTAA TAGAAATTGA
1201 ACCAGAGTTT TCTTATGTTT GCTTGAACAG TTGTGTAAAT CATACAGGAT
1251 TTTGTGGGTA TTGGTTGAAT ATTTGTAAAC CATTCCCTAG CCTACATATT
1301 TATTACTGAA TTAACTTTCC TGATAACCAT TGCATAATTA CATTTTTCTA
1351 TAAAATGAAA GATTATTACA ACAAAAAAAA AAAAAAAAAA AAAAA
```

BLAST Results

Entry HS419346 from database EMBL:
human STS WI-13569.
Score = 2154, P = 8.6e-91, identities = 446/459

Entry HS1292427 from database EMBL:
human STS SHGC-50338.
Score = 1737, P = 7.2e-72, identities = 359/369

Entry HS253344 from database EMBL:
human STS WI-13893.
Score = 1578, P = 1.0e-64, identities = 358/397

Medline entries

No Medline entry

```
                    Peptide information for frame 3
                    -------------------------------

ORF from 156 bp to 917 bp; peptide length: 254
Category: similarity to unknown protein
Classification: no clue
Prosite motifs: RGD (15-18)

    1 MNSRQAWRLF LSQGRGDRWV SRPRGHFSPA LRREFFTTTT KEGYDRRPVD
   51 ITPLEQRKLT FDTHALVQDL ETHGFDKTQA ETIVSALTAL SNVSLDTIYK
  101 EMVTQAQQEI TVQQLMAHLD AIRKDMVILE KSEFANLRAE NEKMKIELDQ
  151 VKQQLMHETS RIRADNKLDI NLERSRVTDM FTDQEKQLME TTTEFTKKDT
  201 QTKSIISETS NKIDAEIASL KTLMESNKLE TIRYLAASVF TCLAIALGFY
  251 RFWK


                              BLASTP hits

No BLASTP hits available

           Alert BLASTP hits for DKFZphtes3_19f19, frame 3

SWISSPROT:YAN8_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME
I., N = 1, Score = 144, P = 8.4e-09

PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces
cerevisiae), N = 1, Score = 138, P = 5.4e-08


>SWISSPROT:YAN8_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME I.
            Length = 211

  HSPs:

 Score = 144 (21.6 bits), Expect = 8.4e-09, P = 8.4e-09
 Identities = 34/121 (28%), Positives = 67/121 (55%)

Query:    70 LETHGFDKTQAETIVSALTALSNVSLDTIYKEMVTQAQQE-ITVQQLMAHLDAIRKDMVI 128
             LE  G+     AETI + +   ++   +L   + K +   +A+QE ++ QQ     L  IRK +
Sbjct:    46 LEQAGYSVKNAETITNLMRTITGEALTELEKNIGFKAKQESVSFQQKRTFLQ-IRKYLET 104

Query:   129 LEKSEFANLRAENEKMKIELDQVKQQLMHETSRIRADNKLDINLERSRVTDMFTDQEKQL 188
             +E++EF   +R  ++K+  E+++ K  L  +      ++ +L++NLE+ R+ D  T +   +
Sbjct:   105 IEENEFDKVRKSSDKLINEIEKTKSSLREDVKTALSEVRLNLNLEKGRMKDAATSRNTNI 164

Query:   189 ME 190
              E
Sbjct:   165 HE 166


             Pedant information for DKFZphtes3_19f19, frame 3
             ------------------------------------------------

                    Report for DKFZphtes3_19f19.3


[LENGTH]         254
[MW]             29505.73
[pI]             6.99
[HOMOL]          PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces cerevisiae)
2e-10
[FUNCAT]         99 unclassified proteins      [S. cerevisiae, YFL046w] 8e-12
[PROSITE]        RGD       1
[KW]             TRANSMEMBRANE  1
[KW]             LOW_COMPLEXITY      5.12 %
[KW]             COILED_COIL        11.02 %

SEQ    MNSRQAWRLFLSQGRGDRWVSRPRGHFSPALRREFFTTTTKEGYDRRPVDITPLEQRKLT
SEG    ............................................................
PRD    ccchhhhhhhhhcccccceeeeccccccchhhhhheeeecccccccccccccchhhhhhcc
COILS  ............................................................
MEM    ............................................................

SEQ    FDTHALVQDLETHGFDKTQAETIVSALTALSNVSLDTIYKEMVTQAQQEITVQQLMAHLD
SEG    ............................................................
PRD    chhhhhhhhhhhccccchhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  ............................................................
```

```
MEM     ............................................................

SEQ     AIRKDMVILEKSEFANLRAENEKMKIELDQVKQQLMHETSRIRADNKLDINLERSRVTDM
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ..............CCCCCCCCCCCCCCCCCCCCCCCCCCCC..................
MEM     ............................................................

SEQ     FTDQEKQLMETTTEFTKKDTQTKSIISETSNKIDAEIASLKTLMESNKLETIRYLAASVF
SEG     ..........xxxxxxxxxxxxx.....................................
PRD     hhhhhhhhhhhhhhhccccccccceeeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ............................................................
MEM     .....................................................MMMMMMM

SEQ     TCLAIALGFYRFWK
SEG     ..............
PRD     hhhhhhhhhhhccc
COILS   ..............
MEM     MMMMMMMMMM....
```

Prosite for DKFZphtes3_19f19.3

```
PS00016      15->18   RGD                    PDOC00016
```

(No Pfam data available for DKFZphtes3_19f19.3)

DKFZphtes3_19j17
----

group: testes derived

DKFZphtes3_19j17 encodes a novel 436 amino acid protein with partial similarity to C.elegans Y40B1A.2 protein.

The novel protein contains two Prosite WW/rsp5/WWP domain signatures.
The WW domain (or rsp5 or WWP domain) has been originally discovered as a short conserved region in a number of unrelated proteins, such as dystrophin, utrophin, vertebrate YAP protein, mouse NEDD-4 and yeast RSP5.
The domain is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans Y40B1A.2

there are two long ORFs in this cDNA according to EST:
HS12146/HS75086/AA923755/MMAA17335 remaining intron at Bp 1506-1733

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp
Poly A stretch at pos. 2740, no polyadenylation signal found

```
   1 ATTCTCAGCC AAATTTTTTT ATTTTTTGCA GAATCAGTGT GCAAGGTGGT
  51 TTATAAGATA ATGGAGTGGT TTTTTTTTGT GTTTAGTGTG ATTTGTTATC
 101 AGGAGTCTTA TTGTAACGCT TAAGCATTAG GTTTTTTGTC TGAGAAACTT
 151 TAAAGAGTAA AGCAGAATTG AAAGTGGAAA TTTTAATTTT GTAAGTTCAT
 201 AAAATTTAAT GATAATACAC CAAAGTTTAT GTTTAAATTA GGGAGTTTAA
 251 GGTTTCAATT CTTTCTCTTT TTTTTTGGGG GGGTGATGTT TTACAGGCAC
 301 TTAAGTATTC ATCGAAGAGT CACCCCAGTA GCGGTGATCA CAGACATGAA
 351 AAGATGCGAG ACGCCGGAGA TCCTTCACCA CCAAATAAAA TGTTGCGGAG
 401 ATCTGATAGT CCTGAAAACA AATACAGTGA CAGCACAGGT CACAGTAAGG
 451 CCAAAAATGT GCATACTCAC AGAGTTAGAG AGAGGGATGG TGGGACCAGT
 501 TACTCTCCAC AAGAAAATTC ACACAACCAC AGTGCTCTTC ATAGTTCAAA
 551 TTCACATTCT TCTAATCCAA GCAATAACCC AAGCAAAACT TCAGATGCAC
 601 CTTATGATTC TGCAGATGAC TGGTCTGAGC ATATTAGCTC TTCTGGGAAA
 651 AAGTACTACT ACAATTGTCG AACAGAAGTT TCACAATGGG AAAAACCAAA
 701 AGAGTGGCTT GAAAGAGAAC AGAGACAAAA AGAAGCAAAC AAGATGGCAG
 751 TCAACAGCTT CCCAAAAGAT AGGGATTACA GAAGAGAGGT GATGCAAGCA
 801 ACAGCCACTA GTGGGTTTGC CAGTGGAATG GAAGACAAGC ATTCCAGTGA
 851 TGCCAGTAGT TTGCTCCCAC AGAATATTTT GTCTCAAACA AGCAGACACA
 901 ATGACAGAGA CTACAGACTG CCAAGAGCAG AGACTCACAG TAGTTCTACG
 951 CCAGTACAGC ACCCCATCAA ACCAGTGGTT CATCCAACTG CTACCCCAAG
1001 CACTGTTCCT TCTAGTCCAT TTACGCTACA GTCTGATCAC CAGCCAAAGA
1051 AATCATTTGA TGCTAATGGA GCATCTACTT TATCAAAACT GCCTACACCC
1101 ACATCTTCTG TCCCTGCACA GAAAACAGAA AGAAAAGAAT CTACATCAGG
1151 AGACAAACCC GTATCACATT CTTGCACAAC TCCTTCCACG TCTTCTGCCT
1201 CTGGACTGAA CCCCACATCT GCACCTCCAA CATCTGCTTC AGCGGTCCCT
1251 GTTTCTCCTG TTCCACAGTC GCCAATACCT CCCTTACTTC AGGACCCAAA
1301 TCTTCTTAGA CAATTGCTTC CTGCTTTGCA AGCCACGCTG CAGCTTAATA
1351 ATTCTAATGT GGACATATCT AAAATAAATG AAGTTCTTAC AGCAGCTGTG
1401 ACACAAGCCT CACTGCAGTC TATAATTCAT AAGTTTCTTA CTGCTGGACC
1451 ATCTGCTTTC AACATAACGT CTCTGATTTC TCAAGCTGCT CAGCTCTCTA
1501 CACAAGATAT CCCTCTTCAT GAAGGTATCC AAATGGAGAG AGATACACAT
1551 AGGAGCAAAT GGGAAGTGAA AGGGTCACTT TGTCAGAAAG CTGATAAACA
1601 GCAGGAATGC CTTGTCTGGA ATGGAAGTAT AATGGTGCAA AGACTCTTGC
1651 AACCCTCTGG CTAGCCTCAT GAGCAGGAGA CTGCGTGGGA TACCTGGGCC
1701 TAAATGTAGA ATAAGAAAGA AGAAATAAGG ATGCCCAGCC ATCTAATCAG
1751 TCTCCGATGT CTTTAACATC TGATGCGTCA TCCCCAAGAT CATATGTTTC
1801 TCCAAGAATA AGCACACCTC AAACTAACAC AGTCCCTATC AAACCTTTGA
1851 TCAGTACTCC TCCTGTTTCA TCACAGCCAA AGGTTAGTAC TCCAGTAGTT
1901 AAGCAAGGAC CAGTGTCACA GTCAGCCACA CAGCAGCCTG TAACTGCTGA
1951 CAAGCAGCAA GGTCATGAAC CTGTCTCTCC TCGAAGTCTT CAGCGCTCAA
2001 GCCAGAGAAG TCCATCACCT GGTCCCAATC ATACTTCTAA TAGTAGTAAT
2051 GCATCAAATG CAACAGTTGT ACCACAGAAT TCTTCTGCCC GATCCACGTG
```

```
2101 TTCATTAACG CCTGCACTAG CAGCACACTT CAGTGAAAAT CTCATAAAAC
2151 ACGTTCAAGG ATGGCCTGCA GATCATGCAG AGAAGCAGGC ATCAAGATTA
2201 CGCGAAGAAG CGCATAACAT GGGAACTATT CACATGTCCG AAATTTGTAC
2251 TGAATTAAAA AATTTAAGAT CTTTAGTCCG AGTATGTGAA ATTCAAGCAA
2301 CTTTGCGAGA GCAAAGGATA CTATTTTTGA GACAACAAAT TAAGGAACTT
2351 GAAAAGCTAA AAAATCAGAA TTCCTTCATG GTGTGAAGAT GTGAATAATT
2401 GCACATGGTT TTGAGAACAG GAACTGTAAA TCTGTTGCCC AATCTTAACA
2451 TTTTTGAGCT GCATTTAAGT AGACTTTGGA CCGTTAAGCT GGGCAAAGGA
2501 AATGACAAGG GGACGGGGTC TGTGAGAGTC AATTCAGGGG AAAGATACAA
2551 GATTGATTTG TAAAACCCTT GAAATGTAGA TTTCTTGTAG ATGTATCCTT
2601 CACGTTGTAA ATATGTTTTG TAGAGTGAAG CCATGGGAAG CCATGTGTAA
2651 CAGAGCTTAG ACATCCAAAA CTAATCAATG CTGAGGTGGC TAAATACCTA
2701 GCCTTTTACA TGTAAACCTG TCTGCAAAAT TAGCTTTTTT AAAAAAAAAA
2751 AAAAAAAAAA AA
```

BLAST Results

Entry AC005876 from database EMBLNEW:
Homo sapiens chromosome 10 clone CIT987SK-1188I5 map 10p11.2-10p12.1, complete sequence.
Score = 2130, P = 0.0e+00, identities = 426/426
12 exons matching Bp 492-2740

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 1757 bp to 2383 bp; peptide length: 209
Category: questionable ORF
Classification: no clue

```
  1 MSLTSDASSP RSYVSPRIST PQTNTVPIKP LISTPPVSSQ PKVSTPVVKQ
 51 GPVSQSATQQ PVTADKQQGH EPVSPRSLQR SSQRSPSPGP NHTSNSSNAS
101 NATVVPQNSS ARSTCSLTPA LAAHFSENLI KHVQGWPADH AEKQASRLRE
151 EAHNMGTIHM SEICTELKNL RSLVRVCEIQ ATLREQRILF LRQQIKELEK
201 LKNQNSFMV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_19j17, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 354 bp to 1661 bp; peptide length: 436
Category: similarity to unknown protein
Classification: unclassified
Prosite motifs: WW_DOMAIN_1 (90-116)
WW_DOMAIN_1 (90-116)

```
  1 MRDAGDPSPP NKMLRRSDSP ENKYSDSTGH SKAKNVHTHR VRERDGGTSY
 51 SPQENSHNHS ALHSSNSHSS NPSNNPSKTS DAPYDSADDW SEHISSSGKK
101 YYYNCRTEVS QWEKPKEWLE REQRQKEANK MAVNSFPKDR DYRREVMQAT
151 ATSGFASGME DKHSSDASSL LPQNILSQTS RHNDRDYRLP RAETHSSSTP
201 VQHPIKPVVH PTATPSTVPS SPFTLQSDHQ PKKSFDANGA STLSKLPTPT
251 SSVPAQKTER KESTSGDKPV SHSCTTPSTS SASGLNPTSA PPTSASAVPV
301 SPVPQSPIPP LLQDPNLLRQ LLPALQATLQ LNNSNVDISK INEVLTAAVT
351 QASLQSIIHK FLTAGPSAFN ITSLISQAAQ LSTQDIPLHE GIQMERDTHR
401 SKWEVKGSLC QKADKQQECL VWNGSIMVQR LLQPSG
```

BLASTP hits

```
No BLASTP hits available

      Alert BLASTP hits for DKFZphtes3_19j17, frame 3

TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid
Y40B1A, N = 1, Score = 144, P = 1.8e-09


>TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A
      Length = 120

 HSPs:

 Score = 144 (21.6 bits), Expect = 1.8e-09, P = 1.8e-09
 Identities = 30/67 (44%), Positives = 43/67 (64%)

Query:  90 WSEHISSSGKKYYYNCRTEVSQWEKPKEW-LEREQRQKEANKMAVNSFPK---DRDYRRE 145
       W+E +SSSGK YYYN +TE+SQW+KP EW E  +++ K VN P+   DR Y
Sbjct:  11 WTEQMSSSGKMYYYNKKTEISQWDKPAEWPAEGGSAERDKPKGGVNEKPRFAEDR-YNEY 69

Query:  146 VMQATATS 153
       + Q +++S
Sbjct:  70 IGQLSSSS 77


      Pedant information for DKFZphtes3_19j17, frame 2
      ----------------------------------------------

           Report for DKFZphtes3_19j17.2


[LENGTH]        209
[MW]            22873.85
[pI]            9.95
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY   13.40 %


SEQ     MSLTSDASSPRSYVSPRISTPQTNTVPIKPLISTPPVSSQPKVSTPVVKQGPVSQSATQQ
SEG     ............................................................
PRD     cccccccccccccccccccccceeeeccccccccccccccccccceeecccccccccccc

SEQ     PVTADKQQGHEPVSPRSLQRSSQRSPSPGPNHTSNSSNASNATVVPQNSSARSTCSLTPA
SEG     .............xxxxxxxxxxxxxx..xxxxxxxxxxxxx..................
PRD     cccccccccccccccccccccccccccccccccccccccceeeecccccccccccchhh

SEQ     LAAHFSENLIKHVQGWPADHAEKQASRLREEAHNMGTIHMSEICTELKNLRSLVRVCEIQ
SEG     ............................................................
PRD     hhhhhhcchhhhhhccccchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhh

SEQ     ATLREQRILFLRQQIKELEKLKNQNSFMV
SEG     .............................
PRD     hhhhhhhhhhhhhhhhhhhhhhhcccccc


(No Prosite data available for DKFZphtes3_19j17.2)

(No Pfam data available for DKFZphtes3_19j17.2)


      Pedant information for DKFZphtes3_19j17, frame 3
      ----------------------------------------------

           Report for DKFZphtes3_19j17.3


[LENGTH]        436
[MW]            47716.62
[pI]            8.71
[HOMOL]         TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A 6e-08

[FUNCAT]        04.05.03 mrna processing (splicing)   [S. cerevisiae, YKL012w] 2e-04
[FUNCAT]        30.10 nuclear organization      [S. cerevisiae, YKL012w] 2e-04
[FUNCAT]        99 unclassified proteins        [S. cerevisiae, YPR152c] 6e-04
[BLOCKS]        BL01159 WW/rsp5/WWP domain proteins
[PROSITE]       WW_DOMAIN_1     2
[PFAM]          WW/rsp5/WWP domain containing proteins
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY   22.48 %
```

```
SEQ   MRDAGDPSPPNKMLRRSDSPENKYSDSTGHSKAKNVHTHRVRERDGGTSYSPQENSHNHS
SEG   ......................................................xxxxxx
PRD   cccccccccccccccccccccccccccccccccccceeeeeecccccccccccccccccc

SEQ   ALHSSNSHSSNPSNNPSKTSDAPYDSADDWSEHISSSGKKYYYNCRTEVSQWEKPKEWLE
SEG   xxxxxxxxxxxxxxxxx...........................................
PRD   cccccccccccccccccccccccccccccceeecccccceeeeeeccccccccccccchhhh

SEQ   REQRQKEANKMAVNSFPKDRDYRREVMQATATSGFASGMEDKHSSDASSLLPQNILSQTS
SEG   ............................................................
PRD   hhhhhhhhhhhhcccccccchhhhhhhhhhccccccccccccccccccccccccccccccc

SEQ   RHNDRDYRLPRAETHSSSTPVQHPIKPVVHPTATPSTVPSSPFTLQSDHQPKKSFDANGA
SEG   ........................xxxxxxxxxxxxxxx.....................
PRD   cccccccccccccccccccccccceeeecccccccccccccccccccccccccccccccc

SEQ   STLSKLPTPTSSVPAQKTERKESTSGDKPVSHSCTTPSTSSASGLNPTSAPPTSASAVPV
SEG   xxxxxxxxxxxx................................xxxxxxxxxxxxxx
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   SPVPQSPIPPLLQDPNLLRQLLPALQATLQLNNSNVDISKINEVLTAAVTQASLQSIIHK
SEG   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx............................
PRD   cccccccccccccccchhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhh

SEQ   FLTAGPSAFNITSLISQAAQLSTQDIPLHEGIQMERDTHRSKWEVKGSLCQKADKQQECL
SEG   ............................................................
PRD   hhcccccceeehhhhhhhhhhhcccccccccccccccccccceeeecccchhhhhhhccee

SEQ   VWNGSIMVQRLLQPSG
SEG   ................
PRD   eeccchhhhhhccccc
```

Prosite for DKFZphtes3_19j17.3

PS01159	90->116	WW_DOMAIN_1	PDOC50020
PS01159	90->116	WW_DOMAIN_1	PDOC50020

Pfam for DKFZphtes3_19j17.3

```
HMM_NAME    WW/rsp5/WWP domain containing proteins

HMM          *LPsGWEeHWDpsGRpWYYWNHETkTTQWEpP*
            + ++W EH++ SG+ YY+N T+ +QWE+P
Query        86 SADDWSEHISSSGKK-YYYNCRTEVSQWEKP  115
```

DKFZphtes3_1c1
----

group: signal transduction

DKFZphtes3_1c1 encodes a novel 632 amino acid putative GTPase-activating protein, related to drosophila rotund transcript and human n-chimaerin.

rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find application in modulating/blocking the response to a cellular receptor.

similarity to GTPase-activating proteins

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3237 bp
Poly A stretch at pos. 3227, no polyadenylation signal found

```
   1 GCGAAGTGAA GGGTGGCCCA GGTGGGGCCA GGCTGACTGA ATGTATCTCC
  51 TAGCTATGGA CTAAATAATA CATGGGGGGA AATAAACAAG TATTCATGAG
 101 GGTGAAAATG TGACCCAGCA GGAAAATTAC AACTATTTTC AATTGACGTT
 151 GAATAGGATG AGTCATGGAA TTTAAGTGAT TTACTGAAGA TTATACTACT
 201 GGTAGATAGA AGAGCTAAAG AAAGATGGAT ACTATGATGC TGAATGTGCG
 251 GAATCTGTTT GAGCAGCTTG TGCGCCGGGT GGAGATTCTC AGTGAAGGAA
 301 ATGAAGTCCA ATTTATCCAG TTGGCGAAGG ACTTTGAGGA TTTCCGTAAA
 351 AAGTGGCAGA GGACTGACCA TGAGCTGGGG AAATACAAGG ATCTTTTGAT
 401 GAAAGCAGAG ACTGAGCGAA GTGCTCTGGA TGTTAAGCTG AAGCATGCAC
 451 GTAATCAGGT GGATGTAGAG ATCAAACGGA GACAGAGAGC TGAGGCTGAC
 501 TGCGAAAAGC TGGAACGACA GATTCAGCTG ATTCGAGAGA TGCTCATGTG
 551 TGACACATCT GGCAGCATTC AACTAAGCGA GGAGCAAAAA TCAGCTCTGG
 601 CTTTTCTCAA CAGAGGCCAA CCATCCAGCA GCAATGCTGG GAACAAAAGA
 651 CTATCAACCA TTGATGAATC TGGTTCCATT TTATCAGATA TCAGCTTTGA
 701 CAAGACTGAT GAATCACTGG ATTGGGACTC TTCTTTGGTG AAGACTTTCA
 751 AACTGAAGAA GAGAGAAAAG AGGCGCTCTA CTAGCCGACA GTTTGTTGAT
 801 GGTCCCCCTG GACCTGTAAA GAAAACTCGT TCCATTGGCT CTGCAGTAGA
 851 CCAGGGGAAT GAATCCATAG TTGCAAAAAC TACAGTGACT GTTCCCAATG
 901 ATGGCGGGCC CATCGAAGCT GTGTCCACTA TTGAGACTGT GCCATATTGG
 951 ACCAGGAGCC GAAGGAAAAC AGGTACTTTA CAACCTTGGA ACAGTGACTC
1001 CACCCTGAAC AGCAGGCAGC TGGAGCCAAG AACTGAGACA GACAGTGTGG
1051 GCACGCCACA GAGTAATGGA GGGATGCGCC TGCATGACTT TGTTTCTAAG
1101 ACGGTTATTA AACCTGAATC CTGTGTTCCA TGTGGAAAGC GGATAAAATT
1151 TGGCAAATTA TCTCTGAAGT GTCGAGACTG TCGTGTGGTC TCTCATCCAG
1201 AATGTCGGGA CCGCTGTCCC CTTCCCTGCA TTCCTACCCT GATAGGAACA
1251 CCTGTCAAGA TTGGAGAGGG AATGCTGGCA GACTTTGTGT CCCAGACTTC
1301 TCCAATGATC CCCTCCATTG TTGTGCATTG TGTAAATGAG ATTGAGCAAA
1351 GAGGTCTGAC TGAGACAGGC CTGTATAGGA TCTCTGGCTG TGACCGCACA
1401 GTAAAAGAGC TGAAAGAGAA ATTCCTCAGA GTGAAAACTG TACCCCTCCT
1451 CAGCAAAGTG GATGATATCC ATGCTATCTG TAGCCTTCTA AAAGACTTTC
1501 TTCGAAACCT CAAAGAACCT CTTCTGACCT TTCGCCTTAA CAGAGCCTTT
1551 ATGGAAGCAG CAGAAATCAC AGATGAAGAC AACAGCATAG CTGCCATGTA
1601 CCAAGCTGTT GGTGAACTGC CCCAGGCCAA CAGGGACACA TTAGCTTTCC
1651 TCATGATTCA CTTGCAGAGA GTGGCTCAGA GTCCACATAC TAAAATGGAT
1701 GTTGCCAATC TGGCTAAAGT CTTTGGCCCT ACAATAGTGG CCCATGCTGT
1751 GCCCAATCCA GACCCAGTGA CAATGTTACA GGACATCAAG CGTCAACCCA
1801 AGGTGGTTGA GCGCCTGCTT TCCTTGCCTC TGGAGTATTG GAGTCAGTTC
1851 ATGATGGTGG AGCAAGAGAA CATTGACCCC CTACATGTCA TTGAAAACTC
1901 AAATGCCTTT TCAACACCAC AGACACCAGA TATTAAAGTG AGTTTACTGG
1951 GACCTGTGAC CACTCCTGAA CATCAGCTTC TCAAGACTCC TTCATCTAGT
2001 TCCCTGTCAC AGAGAGTCCG TTCCACCCTC ACCAAGAACA CTCCTAGATT
2051 TGGGAGCAAA AGCAAGTCTG CCACTAACCT AGGACGACAA GGCAACTTTT
2101 TTGCTTCTCC AATGCTCAAG TGAAGTCACA TCTGCCTGTT ACTTCCCAGC
2151 ATTGACTGAC TATAAGAAAG GACACATCTG TACTCTGCTC TGCAGCCTCC
2201 TGTACTCATT ACTACTTTTA GCATTCTCCA GGCTTTTACT CAAGTTTAAT
2251 TGTGCATGAG GGTTTTATTA AAACTATATA TATCTCCCCT TCCTTCTCCT
2301 CAAGTCACAT AATATCAGCA CTTTGTGCTG GTCATTGTTG GGAGCTTTTA
2351 GATGAGACAT CTTTCCAGGG GTAGAAGGGT TAGTATGGAA TTGGTTGTGA
2401 TTCTTTTTGG GGAAGGGGGT TATTGTTCCT TTGGCTTAAA GCCAAATGCT
2451 GCTCATAGAA TGATCTTTCT CTAGTTTCAT TTAGAACTGA TTTCCGTGAG
2501 ACAATGACAG AAACCCTACC TATCTGATAA GATTAGCTTG TCTCAGGGTG
2551 GGAAGTGGGA GGGCAGGGCA AAGAAAGGAT TAGACCAGAG GATTTAGGAT
```

```
2601 GCCTCCTTCT AAGAACCAGA AGTTCTCATT CCCCATTATG AACTGAGCTA
2651 TAATATGGAG CTTTCATAAA AATGGGATGC ATTGAGGACA GAACTAGTGA
2701 TGGGAGTATG CGTAGCTTTG ATTTGGATGA TTAGGTCTTT AATAGTGTTG
2751 AGTGGCACAA CCTTGTAAAT GTGAAAGTAC AACTCGTATT TATCTCTGAT
2801 GTGCCGCTGG CTGAACTTTG GGTTCATTTG GGGTCAAAGC CAGTTTTTCT
2851 TTTAAAATTG AATTCATTCT GATGCTTGGC CCCCATACCC CCAACCTTGT
2901 CCAGTGGAGC CCAACTTCTA AAGGTCAATA TATCATCCTT TGGCATCCCA
2951 ACTAACAATA AAGAGTAGGC TATAAGGGAA GATTGTCAAT ATTTTGTGGT
3001 AAGAAAAGCT ACAGTCATTT TTTCTTTGCA CTTTGGATGC TGAAATTTTT
3051 CCCATGGAAC ATAGCCACAT CTAGATAGAT GTGAGCTTTT TCTTCTGTTA
3101 AAATTATTCT TAATGTCTGT AAAAACGATT TTCTTCTGTA GAATGTTTGA
3151 CTTCGTATTG ACCCTTATCT GTAAAACACC TATTTGGGAT AATATTTGGA
3201 AAAAAAGTAA ATAGCTTTTT CAAAATGAAA AAAAAAA
```

BLAST Results

Entry U82984 from database EMBLEST:
Homo sapiens DRES 56 mRNA sequence.
Score = 8775, P = 0.0e+00, identities = 1757/1758
matches 3' end

Medline entries

93074974:
Developmental regulation and neuronal expression of the mRNA of rat n-chimaerin, a
p21rac GAP:cDNA sequence.

93024458:
A Drosophila rotund transcript expressed during spermatogenesis and imaginal disc
morphogenesis encodes a protein which is similar to human Rac GTPase-activating
(racGAP) proteins.

Peptide information for frame 3

ORF from 225 bp to 2120 bp; peptide length: 632
Category: similarity to known protein

```
  1 MDTMMLNVRN LFEQLVRRVE ILSEGNEVQF IQLAKDFEDF RKKWQRTDHE
 51 LGKYKDLLMK AETERSALDV KLKHARNQVD VEIKRRQRAE ADCEKLERQI
101 QLIREMLMCD TSGSIQLSEE QKSALAFLNR GQPSSSNAGN KRLSTIDESG
151 SILSDISFDK TDESLDWDSS LVKTFKLKKR EKRRSTSRQF VDGPPGPVKK
201 TRSIGSAVDQ GNESIVAKTT VTVPNDGGPI EAVSTIETVP YWTRSRRKTG
251 TLQPWNSDST LNSRQLEPRT ETDSVGTPQS NGGMRLHDFV SKTVIKPESC
301 VPCGKRIKFG KLSLKCRDCR VVSHPECRDR CPLPCIPTLI GTPVKIGEGM
351 LADFVSQTSP MIPSIVVHCV NEIEQRGLTE TGLYRISGCD RTVKELKEKF
401 LRVKTVPLLS KVDDIHAICS LLKDFLRNLK EPLLTFRLNR AFMEAAEITD
451 EDNSIAAMYQ AVGELPQANR DTLAFLMIHL QRVAQSPHTK MDVANLAKVF
501 GPTIVAHAVP NPDPVTMLQD IKRQPKVVER LLSLPLEYWS QFMMVEQENI
551 DPLHVIENSN AFSTPQTPDI KVSLLGPVTT PEHQLLKTPS SSSLSQRVRS
601 TLTKNTPRFG SKSKSATNLG RQGNFFASPM LK
```

BLASTP hits

Entry CEK08E3_4 from database TREMBLNEW:
gene: "K08E3.6"; Caenorhabditis elegans cosmid K08E3
Score = 452, P = 2.6e-48, identities = 126/377, positives = 189/377

Entry A48122 from database PIR:
GTPase-activating protein Rac homolog, splice form clone pcl.7 - fruit fly (Drosophila melanogaster) (fragment)
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry B48122 from database PIR:
GTPase-activating protein Rac homolog, splice form clone pcl.7d - fruit fly (Drosophila melanogaster)
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

```
Entry DM22539_1 from database TREMBL:
gene: "rotund"; product: "rnracGAP"; Drosophila melanogaster rnracGAP
(rotund) gene, complete cds.
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry S29128 from database PIR:
N-chimerin - rat
Score = 336, P = 8.8e-30, identities = 86/253, positives = 128/253


            Alert BLASTP hits for DKFZphtes3_1c1, frame 3

No Alert BLASTP hits found

            Pedant information for DKFZphtes3_1c1, frame 3
            ------------------------------------------

                     Report for DKFZphtes3_1c1.3


[LENGTH]        632
[MW]            71026.84
[pI]            9.08
[HOMOL]         PIR:B48122 GTPase-activating protein Rac homolog, splice form clone pcl.7d -
fruit fly (Drosophila melanogaster) 2e-46
[FUNCAT]        10.99 other signal-transduction activities   [S. cerevisiae, YBR260c] 3e-12
[FUNCAT]        03.22 cell cycle control and mitosis [S. cerevisiae, YER155c] 2e-11
[FUNCAT]        30.03 organization of cytoplasm      [S. cerevisiae, YER155c] 2e-11
[FUNCAT]        03.04 budding, cell polarity and filament formation [S. cerevisiae, YER155c]
2e-11
[FUNCAT]        03.10 sporulation and germination    [S. cerevisiae, YDL240w] 3e-09
[FUNCAT]        30.04 organization of cytoskeleton   [S. cerevisiae, YOR134w] 4e-09
[FUNCAT]        06.10 assembly of protein complexes  [S. cerevisiae, YOR134w] 4e-09
[FUNCAT]        03.07 pheromone response, mating-type determination, sex-specific proteins
      [S. cerevisiae, YOR127w] 5e-09
[FUNCAT]        09.04 biogenesis of cytoskeleton     [S. cerevisiae, YPL115c] 3e-08
[FUNCAT]        10.02.09 regulation of g-protein activity   [S. cerevisiae, YPL115c] 3e-08
[BLOCKS]        BL00479B Phorbol esters / diacylglycerol binding domain proteins
[BLOCKS]        BL00479A Phorbol esters / diacylglycerol binding domain proteins
[SCOP]          d1pbwa_ 1.83.1.1.2 p85 alpha subunit RhoGAP domain (human (Hom 1e-55
[SCOP]          d1rgp__ 1.83.1.1.1 p50 RhoGAP domain (human (Homo sapiens) 1e-49
[PIRKW]         breakpoint cluster region 1e-19
[PIRKW]         transmembrane protein 7e-08
[PIRKW]         brain 3e-22
[PIRKW]         alternative splicing 1e-19
[PIRKW]         P-loop 2e-25
[SUPFAM]        CDC24 homology 3e-22
[SUPFAM]        bcr protein 3e-22
[SUPFAM]        myosin motor domain homology 2e-25
[SUPFAM]        pleckstrin repeat homology 4e-10
[SUPFAM]        LIM metal-binding repeat homology 2e-09
[SUPFAM]        protein kinase C zinc-binding repeat homology 5e-29
[PROSITE]       MYRISTYL        6
[PROSITE]       AMIDATION       1
[PROSITE]       CAMP_PHOSPHO_SITE       3
[PROSITE]       CK2_PHOSPHO_SITE        13
[PROSITE]       TYR_PHOSPHO_SITE        2
[PROSITE]       PKC_PHOSPHO_SITE        9
[PROSITE]       ASN_GLYCOSYLATION       1
[PROSITE]       DAG_PE_BINDING_DOMAIN   1
[PFAM]          Phorbol esters / diacylglycerol binding domain
[KW]            Irregular
[KW]            3D
[KW]            LOW_COMPLEXITY      2.22 %
[KW]            COILED_COIL         8.54 %


SEQ     MDTMMLNVRNLFEQLVRRVEILSEGNEVQFIQLAKDFEDFRKKWQRTDHELGKYKDLLMK
SEG     ............................................................
COILS   ................................................CCCCCCCCCCCC
1rgp-   ............................................................

SEQ     AETERSALDVKLKHARNQVDVEIKRRQRAEADCEKLERQIQLIREMLMCDTSGSIQLSEE
SEG     ............................................................
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC..................
1rgp-   ............................................................

SEQ     QKSALAFLNRGQPSSSNAGNKRLSTIDESGSILSDISFDKTDESLDWDSSLVKTFKLKKR
SEG     ............................................................
COILS   ............................................................
```

```
1rgp-   ............................................................

SEQ     EKRRSTSRQFVDGPPGPVKKTRSIGSAVDQGNESIVAKTTVTVPNDGGPIEAVSTIETVP
SEG     ............................................................
COILS   ............................................................
1rgp-   ............................................................

SEQ     YWTRSRRKTGTLQPWNSDSTLNSRQLEPRTETDSVGTPQSNGGMRLHDFVSKTVIKPESC
SEG     ............................................................
COILS   ............................................................
1rgp-   ............................................................

SEQ     VPCGKRIKFGKLSLKCRDCRVVSHPECRDRCPLPCIPTLIGTPVKIGEGMLADFVSQTSP
SEG     ............................................................
COILS   ............................................................
1rgp-   ............................................................

SEQ     MIPSIVVHCVNEIEQRGLTETGLYRISGCDRTVKELKEKFLRVKTVPLLSKVDDIHAICS
SEG     ............................................................
COILS   ............................................................
1rgp-   .CCHHHHHHHHHHHHHHTTTTTTTTTCCCHHHHHHHHHHHHHCCCCCG-GGCCCCHHHHH

SEQ     LLKDFLRNLKEPLLTFRLNRAFMEAAEITDEDNSIAAMYQAVGELPQANRDTLAFLMIHL
SEG     ............................................................
COILS   ............................................................
1rgp-   HHHHHHHTTTTTTTGGGHHHHHHTTTT-CGGGHHHHHHHHHHHHCCHHHHHHHHHHHHHH

SEQ     QRVAQSPHTKMDVANLAKVFGPTIVAHAVPNPDPVTMLQDIKRQPKVVERLLSLPLEYWS
SEG     ............................................................
COILS   ............................................................
1rgp-   HHHHHHHHHCCCHHHHHHHHGGGCC...................................

SEQ     QFMMVEQENIDPLHVIENSNAFSTPQTPDIKVSLLGPVTTPEHQLLKTPSSSSLSQRVRS
SEG     .................................................xxxxxxxxxxx
COILS   ............................................................
1rgp-   ............................................................

SEQ     TLTKNTPRFGSKSKSATNLGRQGNFFASPMLK
SEG     xxx.............................
COILS   ................................
1rgp-   ................................
```

Prosite for DKFZphtes3_1c1.3

PS00001	212->216	ASN_GLYCOSYLATION	PDOC00001
PS00004	141->145	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	182->186	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	246->250	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	174->177	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	245->248	PKC_PHOSPHO_SITE	PDOC00005
PS00005	313->316	PKC_PHOSPHO_SITE	PDOC00005
PS00005	392->395	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	595->598	PKC_PHOSPHO_SITE	PDOC00005
PS00005	606->609	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	66->70	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	270->274	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	387->391	CK2_PHOSPHO_SITE	PDOC00006
PS00006	392->396	CK2_PHOSPHO_SITE	PDOC00006
PS00006	410->414	CK2_PHOSPHO_SITE	PDOC00006
PS00006	449->453	CK2_PHOSPHO_SITE	PDOC00006
PS00006	489->493	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00007	46->55	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->385	TYR_PHOSPHO_SITE	PDOC00007
PS00008	131->137	MYRISTYL	PDOC00008
PS00008	150->156	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	377->383	MYRISTYL	PDOC00008
PS00008	388->394	MYRISTYL	PDOC00008
PS00008	623->629	MYRISTYL	PDOC00008
PS00009	303->307	AMIDATION	PDOC00009

```
PS00479     287->336    DAG_PE_BINDING_DOMAIN  PDOC00379


                        Pfam for DKFZphtes3_1cl.3

HMM_NAME       Phorbol esters / diacylglycerol binding domain

HMM               *HrFmrHTFrqPTWCDHCgeFIWGWgKQGYQCQnCgMNCHKRCHelVPmm
                   H+F+ +T + P +C  CG +I  +GK  ++C +C+++ H +C+ + P
Query          287 HDFVSKTVIKPESCVPCGKRI-KFGKLSLKCRDCRVVSHPECRDRCPLP   334

HMM               C*
                  C
Query          335 C    335
```

DKFZphtes3_1g13

group: intracellular transport and trafficking

DKFZp DKFZphtes3_1g13 encodes a novel 1007 amino acid protein with similarity to human 256 kD golgin.

The new protein contains 7 leucine zippers and seems to be involved in protein-protein-interaction in the golgi apparatus. The very similar rat cp151 shows haploid-specific transcription in mus musculus testis.

The new protein can find application in modulating protein traffic in the golgi apparatus, especially in human haploid germ cells.

similarity to 256 kD golgi, strong similarity to rat "cp151"

21 exons encoded on AC004682
EST from a testis library, two mouse ESTs of a testis cDNA library,
rat cp151 shows haploid-specific transcription!
testis or haploid-specific transcription

Sequenced by DKFZ

Locus: map="16q22.2"

Insert length: 3405 bp
Poly A stretch at pos. 3394, polyadenylation signal at pos. 3373

```
   1 GGGATAGGGG ATGTGGTTTG TTACAAAGGA TGAGTATTTT GATAGCTTCT
  51 CATTCCTTGA ACTATTCTGC AGGTTTATAA CAAAGCTCAG AAAATACTAA
 101 AGGTTAAAGG AGAATTGAGA GCTGCCAAGG AAATGAAAGA TGAGGCGGGG
 151 GAGAGAGACA GAGAAGTGAG CAGCCTGAAC AGCAAGCTGT TAAGCCTGCA
 201 ACTTGACATC AAGAATCTGC ACGATGTCTG CAAGAGACAG AGGAAGACCT
 251 TGCAGGACAA TCAGCTCTGC ATGGAGGAGG CAATGAACAG CAGCCACGAC
 301 AAGAAGCAAG CACAGGCATT AGCATTCGAG GAGTCAGAGG TGGAATTTGG
 351 GTCCAGTAAA CAGTGTCATC TGAGACAACT CCAGCAACTG AAGAAAAAAT
 401 TGCTGGTCCT TCAACAAGAA CTGGAGTTTC ACACAGAGGA GTTGCAGACT
 451 TCTTACTATT CTCTCCGCCA GTATCAGTCC ATCCTAGAGA AGCAGACTTC
 501 CGACCTGGTT CTTCTGCACC ATCACTGCAA ACTGAAAGAA GATGAGGTGA
 551 TTCTCTATGA GGAGGAAATG GGAAATCACA ACGAGAACAC AGGGGAGAAG
 601 CTCCATTTGG CGCAGGAGCA ACTCGCCTTG GCCGGGGACA AGATCGCCTC
 651 TCTAGAGAGG AGCTTAAACC TCTACAGGGA TAAATACCAG TCTTCCCTGA
 701 GCAACATCGA GTTACTAGAA TGCCAAGTGA AGATGTTGCA GGGGGAACTC
 751 GGCGGGATCA TGGGTCAGGA GCCTGAGAAC AAGGGTGATC ATTCAAAGGT
 801 ACGGATATAC ACTTCTCCTT GCATGATTCA AGAGCATCAG GAGACTCAGA
 851 AACGACTGTC TGAAGTCTGG CAAAAGGTCT CTCAACAGGA TGATCTCATT
 901 CAAGAACTTC GAAATAAGCT GGCCTGCAGT AACGCTTTGG TTCTGGAGCG
 951 TGAAAAGGCT TTGATAAAAC TACAAGCCGA TTTTGCTTCC TGTACAGCCA
1001 CCCACAGATA CCCTCCTAGC TCCTCAGAAG AGTGTGAAGA CATCAAAAAG
1051 ATACTGAAGC ACTTGCAGGA GCAGAAAGAC AGCCAGTGCC TGCATGTGGA
1101 GGAGTACCAG AACCTGGTGA AGGATCTGCG CGTGGAACTA GAGGCCGTGT
1151 CGGAACAGAA GAGAAACATC ATGAAGGACA TGATGAAGCT GGAGCTGGAC
1201 CTGCACGGAC TGCGGGAGGA GACATCTGCC CACATTGAGA GGAAGGATAA
1251 GGACATCACC ATCCTGCAGT GCCGGCTGCA GGAGCTGCAG CTGGAGTTCA
1301 CCGAGACCCA AAAGCTCACT TTGAAGAAAG ACAAGTTCCT CCAAGAGAAA
1351 GATGAGATGC TGCAAGAGCT GGAGAAGAAA CTGACACAGG TTCAGAACAG
1401 CCTCCTGAAA AAGGAGAAGG AGCTGGAGAA GCAGCAGTGC ATGGCCACAG
1451 AACTTGAAAT GACAGTCAAG GAGGCTAAGC AGGACAAGTC CAAGGAGGCG
1501 GAGTGCAAGG CCCTGCAGGC TGAGGTCCAG AAGCTGAAGA ACAGTCTCGA
1551 AGAGGCCAAG CAGCAGGAGA GGCTGGCTGC TCAGCAAGCA GCCCAGTGCA
1601 AAGAAGAGGC TGCACTGGCA GGCTGTCACC TGGAGGACAC CCAGAGGAAA
1651 CTGCAGAAGG GTCTCCTCCT GGACAAGCAG AAGGCAGACA CCATCCAGGA
1701 ACTACAGAGA GAACTTCAGA TGCTGCAGAA GGAGTCCTCG ATGGCTGAGA
1751 AGGAACAAAC CTCCAACAGA AAACGGGTGG AGGAGCTGTC ATTAGAACTC
1801 TCTGAAGCCC TGAGGAAGCT TGAAAATTCA GACAAGGAAA AGAGGCAGCT
1851 TCAGAAGACA GTGGCTGAGC AGGATATGAA AATGAATGAC ATGCTTGATC
1901 GTATCAAGCA CCAGCACAGG GAGCAAGGCT CCATCAAATG CAAGTTAGAA
1951 GAAGATCTTC AGGAGGCCAC AAAGCTTCTG GAGGACAAAC GGGAGCAGTT
2001 GAAGAAGAGC AAAGAGCATG AGAAGCTGAT GGAGGGAGAA CTTGAAGCTT
2051 TGCGGCAGGA ATTTAAAAAG AAAGACAAGA CGTTGAAAGA GAATTCCAGA
2101 AAGTTGGAGG AAGAAAATGA GAATCTCCGA GCAGAGCTAC AGTGTTGTTC
2151 TACACAACTG GAATCCTCTC TCAACAAATA CAACACCAGC CAGCAAGTCA
2201 TCCAAGACTT GAATAAAGAG ATAGCCCTTC AGAAGGAGTC CTTAATGAGC
2251 CTGCAGGCCC AGCTGGACAA AGCTCTGCAG AAGGAGAAGC ACTATCTCCA
2301 GACTACCATC ACCAAAGAAG CCTATGATGC ATTATCCCGG AAGTCAGCCG
2351 CCTGCCAGGA TGACCTGACA CAAGCCCTCG AGAAGCTCAA TCACGTGACC
2401 TCAGAGACAA AGAGCCTGCA GCAAAGCTTG ACACAGACCC AAGAGAAGAA
```

```
2451 AGCTCAGCTG GAAGAGGAAA TCATTGCTTA TGAGGAAAGG ATGAAAAAGC
2501 TCAATACGGA ATTAAGAAAA CTGCGGGGCT TCCACCAGGA GAGTGAGCTG
2551 GAGGTGCACG CCTTTGACAA GAAGCTAGAG GAGATGAGCT GCCAGGTGCT
2601 GCAGTGGCAG AAGCAACACC AGAATGACCT CAAGATGCTG GCAGCCAAAG
2651 AGGAGCAGCT CAGGGAGTTC CAGGAGGAGA TGGCCGCCTT AAAAGAGAAC
2701 CTCCTTGAGG ACGATAAGGA GCCCTGCTGC CTGCCCCAGT GGTCTGTGCC
2751 CAAAGACACC TGTAGGCTCT ACCGAGGGAA TGATCAGATT ATGACCAACT
2801 TGGAGCAATG GGCAAAACAG CAGAAGGTCG CCAATGAGAA ACTAGGAAAC
2851 CAGCTCCGAG AGCAGGTGAA CTACATTGCC AAGCTGAGTG GCGAAAAGGA
2901 CCACCTCCAC AGTGTAATGG TCCACTTGCA GCAGGAAAAC AAGAAGCTGA
2951 AGAAGGAGAT AGAAGAGAAG AAGATGAAAG CCGAGAACAC AAGGCTATGC
3001 ACCAAAGCCC TAGGCCCGAG CAGAACGGAG TCCACACAGA GGGAGAAAGT
3051 GTGCGGCACC TTGGGCTGGA AGGGGTTGCC CCAGGATATG GGTCAAAGAA
3101 TGGACCTCAC CAAGTACATC GGGATGCCCC ACTGCCCGGG TTCCTCATAC
3151 TGCTAGAATC CACATCTAGC CCTGAGCAGC ATTTCCACGG GTGTTTCTTC
3201 AGAGGACAGT GAGTTCCCAG CCCTCCCTCT CTCTTGACCT GGATCAGCTC
3251 TTACAGGAGT ATATCACGGT CCCAGCCTAT TTTGCAAGAC ACTAACTTTT
3301 GTTGAGTTTT GTCCACTTCC TGCCATGGAG TGAGCTTTAG AACCATACTA
3351 CCATCTCCAG GCCCAAACTC TGAAATAAAG ACATGAGCAT GAGCAAAAAA
3401 AAAAA
```

BLAST Results

Entry AC004682 from database EMBLNEW:
Homo sapiens Chromosome 16 BAC clone CIT987SK-A-259H10, complete sequence.
Score = 1291, P = 0.0e+00, identities = 265/272

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 133 bp to 3153 bp; peptide length: 1007
Category: similarity to known protein
Prosite motifs: LEUCINE_ZIPPER (83-105)
LEUCINE_ZIPPER (90-112)
LEUCINE_ZIPPER (97-119)
LEUCINE_ZIPPER (104-126)
LEUCINE_ZIPPER (403-425)
LEUCINE_ZIPPER (410-432)
LEUCINE_ZIPPER (918-940)

```
   1 MKDEAGERDR EVSSLNSKLL SLQLDIKNLH DVCKRQRKTL QDNQLCMEEA
  51 MNSSHDKKQA QALAFEESEV EFGSSKQCHL RQLQQLKKKL LVLQQELEFH
 101 TEELQTSYYS LRQYQSILEK QTSDLVLLHH HCKLKEDEVI LYEEEMGNHN
 151 ENTGEKLHLA QEQLALAGDK IASLERSLNL YRDKYQSSLS NIELLECQVK
 201 MLQGELGGIM GQEPENKGDH SKVRIYTSPC MIQEHQETQK RLSEVWQKVS
 251 QQDDLIQELR NKLACSNALV LEREKALIKL QADFASCTAT HRYPPSSSEE
 301 CEDIKKILKH LQEQKDSQCL HVEEYQNLVK DLRVELEAVS EQKRNIMKDM
 351 MKLELDLHGL REETSAHIER KDKDITILQC RLQELQLEFT ETQKLTLKKD
 401 KFLQEKDEML QELEKKLTQV QNSLLKKEKE LEKQQCMATE LEMTVKEAKQ
 451 DKSKEAECKA LQAEVQKLKN SLEEAKQQER LAAQQAAQCK EEAALAGCHL
 501 EDTQRKLQKG LLLDKQKADT IQELQRELQM LQKESSMAEK EQTSNRKRVE
 551 ELSLELSEAL RKLENSDKEK RQLQKTVAEQ DMKMNDMLDR IKHQHREQGS
 601 IKCKLEEDLQ EATKLLEDKR EQLKKSKEHE KLMEGELEAL RQEFKKKDKT
 651 LKENSRKLEE ENENLRAELQ CCSTQLESSL NKYNTSQQVI QDLNKEIALQ
 701 KESLMSLQAQ LDKALQKEKH YLQTTITKEA YDALSRKSAA CQDDLTQALE
 751 KLNHVTSETK SLQQSLTQTQ EKKAQLEEEI IAYEERMKKL NTELRKLRGF
 801 HQESELEVHA FDKKLEEMSC QVLQWQKQHQ NDLKMLAAKE EQLREFQEEM
 851 AALKENLLED DKEPCCLPQW SVPKDTCRLY RGNDQIMTNL EQWAKQQKVA
 901 NEKLGNQLRE QVNYIAKLSG EKDHLHSVMV HLQQENKKLK KEIEEKKMKA
 951 ENTRLCTKAL GPSRTESTQR EKVCGTLGWK GLPQDMGQRM DLTKYIGMPH
1001 CPGSSYC
```

BLASTP hits

Entry HS417401_1 from database TREMBL:
product: "trans-Golgi p230"; Human trans-Golgi p230 mRNA, complete

cds.
Score = 411, P = 3.9e-34, identities = 212/862, positives = 420/862

Entry SCINTANA_1 from database TREMBL:
Saccharomyces cerevisiae integrin analogue gene, complete cds.
Score = 404, P = 6.2e-34, identities = 199/897, positives = 423/897

Entry HS6802_2 from database TREMBL:
gene: "MYH9"; product: "dJ6802.2"; Homo sapiens DNA sequence from PAC 6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain, ESTs, CA repeat, STS and GSS.
Score = 404, P = 1.9e-33, identities = 231/1028, positives = 469/1028

Entry AF092090_1 from database TREMBL:
product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.
Score = 2523, P = 3.0e-262, identities = 506/733, positives = 611/733

Alert BLASTP hits for DKFZphtes3_1g13, frame 1

TREMBL:HSGOLGIN_1 product: "256 kD golgin"; H.sapiens mRNA for golgin, N = 1, Score = 411, P = 4.4e-34

TREMBL:HS417401_1 product: "trans-Golgi p230"; Human trans-Golgi p230 mRNA, complete cds., N = 1, Score = 411, P = 4.5e-34

TREMBL:SCINTANA_1 Saccharomyces cerevisiae integrin analogue gene, complete cds., N = 1, Score = 404, P = 7.1e-34

>TREMBL:HSGOLGIN_1 product: "256 kD golgin"; H.sapiens mRNA for golgin
Length = 2,185

HSPs:

Score = 411 (61.7 bits), Expect = 4.4e-34, P = 4.4e-34
Identities = 212/816 (25%), Positives = 420/816 (51%)

```
Query:   145 EMGNHNEN-TGEKLHLAQEQLALAGDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQ 203
             +M +  E+ G    L +EQL      ++  +ERSL+ YR KY   ++  ++L+ + K LQ
Sbjct:   119 DMDSEAEDLVGNSDSLNKEQLI---QRLRRMERSLSSYRGKYSELVTAYQMLQREKKKLQ 175

Query:   204 GELGGIMGQEPENKGDHSKVRIYTSPCMIQEHQETQKRLSEVWQ-KVSQQDDLIQELRNK 262
             G    I+ Q      D S  RI       +Q  Q+ +K L E +    + ++D  I  L+ +
Sbjct:   176 G----ILSQSQ----DKSLRRIAELREELQMDQQAKKHLQEEFDASLEEKDQYISVLQTQ 227

Query:   263 LAC------SNALVLEREKALIKLQADFASCTATHRYPPSSSEEC-ED--IKKILKHLQE 313
             ++       +  + ++  K L +L+   A          P S  E   ED    K L+ LQ+
Sbjct:   228 VSLLKQRLRNGPMNVDVLKPLPQLEPQ-AEVFTKEENPESDGEPVVEDGTSVKTLETLQQ 286

Query:   314 QKDSQ------CLH-VEEYQNLVKDLRVELEAVSEQKRNIMKDMMKLELDLHGLREETSA 366
              +   Q       C    ++ ++       L  E EA+ EQ      ++++ K++ DLH  + E+T
Sbjct:   287 RVKRQENLLKRCKETIQSHKEQCTLLTSEKEALQEQLDERLQELEKIK-DLH-MAEKTKL 344

Query:   367 HIERKDKDITILQCRLQELQLEFTETQKLTLKKDKFLQEKDEMLQELEKKLTQV--QNSL 424
               + +D     I Q    Q+  +    ET++   +  + L+ K+E + +L  ++ Q+  Q
Sbjct:   345 ITQLRDAKNLIEQLE-QDKGMVIAETKR---QMHETLEMKEEEIAQLRSRIKQMTTQGEE 400

Query:   425 LKKEKELEKQQCMATELEMTVKEAKQDKSKEAECKALQAEVQKLKNSLEEAKQQERLAAQ 484
             L+++KE + ++       ELE   A+   K++EA  K L+AE+ +    ++E+   ++ER++ Q
Sbjct:   401 LREQKE-KSERAAFEELEKALSTAQ--KTEEARRK-LKAEMDEQIKTIEKTSEEERISLQ 456

Query:   485 QA-AQCKEEAA-LAGCHLEDTQRKLQKGLLLDKQKADTIQELQRELQMLQKESSMAEKEQ 542
             Q  ++ K+E     +     E+    KLQK  L +K+ A   QEL ++LQ  ++E      E+ +
Sbjct:   457 QELSRVKQEVVDVMKKSSEEQIAKLQK--LHEKELARKEQELTKKLQTRERE--FQEQMK 512

Query:   543 TSNRKRVEELSLELSEALRKLENSDKEKRQLQKT--VAEQDMKMNDMLDRIKHQHREQGS 600
               +  K    E  L++S+   + E+    E+ +LQK    + E + K+ D+      +
Sbjct:   513 VALEKSQSEY-LKISQEKEQQESLALEELELQKKAILTESENKLRDLQQEAETYRTRILE 571

Query:   601 IKCKLEEDLQEATKLLED-----KREQLKKSKEHEKLMEG---ELEALR-QEFKKKDKTL 651
             ++  LE+ LQE     +D     + E+ K +KE    ++E      ELE+L+ Q+      + L
Sbjct:   572 LESSLEKSLQENKNQSKDLAVHLEAEKNKHNKEITVMVEKHKTELESLKHQQDALWTEKL 631

Query:   652 KENSRKLEEENENLRAELQCCSTQLESSL-NKYNTSQQVIQDLNKE----IALQKESLMS 706
             +    ++ + E E LR +   C    + E+ L +K    Q   I+++N++    + +++  L S
Sbjct:   632 QVLKQQYQTEMEKLREK---CEQEKETLLKDKEIIFQAHIEEMNEKTLEKLDVKQTELES 688

Query:   707 LQAQLDKALQKEKHYLQT--TITKEAYDALSRKSAACQDDLTQALEKLNHVTSETKSLQQ 764
             L ++L + L K +H L+     ++ K+   D + ++    A  D+  Q           V S  K +
```

```
Sbjct:   689 LSSELSEVL-KARHKLEEELSVLKDQTDKMKQELEAKMDE--QKNHHQQQVDSIIKEHEV 745

Query:   765 SLTQTQEKKAQLEEEIIAYEERMKKLNTELRKLRGFHQESELEVHAFDKKLEEMSCQVLQ 824
             S+ +T+  KA L+++I   E  +K+ +  L++ +   +  E ++   + +L++ S ++
Sbjct:   746 SIQRTE--KA-LKDQINQLELLLKERDKHLKEHQAHVENLEADIKRSEGELQQASAKLDV 802

Query:   825 WQKQHQNDLKMLAAKEEQLREFQEEMAALKENLLEDDKEPCCLPQW-----SVPKDTC-R 878
             +Q  +Q+       A  EQ + ++E++A L++ LL+ + E   L +       +  KD C
Sbjct:   803 FQS-YQS------ATHEQTKAYEEQLAQLQQKLLDLETERILLTKQVAEVEAQKKDVCTE 855

Query:   879 LYRGNDQIMTNLEQWAKQQKVANEKLGNQLREQVNYIAKLS-GEKDHLHSVMVHLQQENK 937
             L    Q+    ++Q  KQ      +K+ +      QV Y +KL  G K+   +  + +++EN
Sbjct:   856 LDAHKIQVQDLMQQLEKQNSEMEQKVKSLT--QV-YESKLEDGNKEQEQTKQILVEKENM 912

Query:   938 KLK-KEIEEKKMKAENTRLCTK 958
             L+ +E ++K+++     +L  K
Sbjct:   913 ILQMREGQKKEIEILTQKLSAK 934

 Score = 338 (50.7 bits), Expect = 3.1e-26, P = 3.1e-26
 Identities = 216/953 (22%), Positives = 468/953 (49%)

Query:     2 KDEAGERDRE--VSSLNS-KLL-SLQLDIKNLHDVCKRQRKTLQDN-QLCM-----EEAM 51
             K+E  E D E  V     S K L +LQ  +K   ++ KR ++T+Q + + C       +EA+
Sbjct:   260 KEENPESDGEPVVEDGTSVKTLETLQQRVKRQENLLKRCKETIQSHKEQCTLLTSEKEAL 319

Query:    52 NSSHDKKQAQALAFEESEVEFGSSKQCHLRQ----LQQLK--KKLLVLQQELEFHTEELQ 105
                D++  +      ++  +   +      LR      ++QL+  K +++ + + + H E L+
Sbjct:   320 QEQLDERLQELEKIKDLHMAEKTKLITQLRDAKNLIEQLEQDKGMVIAETKRQMH-ETLE 378

Query:   106 TSYYSLRQYQSILEKQTSDLVLLHHHCKLKEDEVILYEEEMGNHNENTGEKLHLAQEQL- 164
                + Q +S +++ T+      L      K K +        E E    +T +K    A+ +L
Sbjct:   379 MKEEEIAQLRSRIKQMTTQGEELREQ-KEKSERAAFEELEKAL---STAQKTEEARRKLK 434

Query:   165 ALAGDKIASLERSLNLYRDKYQSSLSNI--ELLECQVKMLQGELGGIMGQEPENKGDHSK 222
             A    ++I ++E++     R    Q LS + E+++    K  + ++  + Q+     K    K
Sbjct:   435 AEMDEQIKTIEKTSEEERISLQQELSRVKQEVVDVMKKSSEEQIAKL--QKLHEKELARK 492

Query:   223 VRIYTSPCMIQEHQETQKRLSEVWQKVSQQDDLIQELRNKLACSNALVLEREKALIKLQA 282
               +  T        +E +E Q+++     +K SQ + L   ++  +      +L LE      ++LQ
Sbjct:   493 EQELTKKLQTRE-REFQEQMKVALEK-SQSEYL--KISQEKEQQESLALEE----LELQK 544

Query:   283 DFASCTATHRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAV-SE 341
               A  T +           +E E  +  +  L+  +        ++E +N  KDL V LEA ++
Sbjct:   545 K-AILTESENKLRDLQQEAETYRTRILELESSLEKS---LQENKNQSKDLAVHLEAEKNK 600

Query:   342 QKRNIMKDMMKLELDLHGLREETSAHIERKDKDITI-LQCRLQELQLEFTETQKLTLKKD 400
                + I    + K + +L   L+ +  A     K + +     Q  +++L+ E  E +K TL KD
Sbjct:   601 HNKEITVMVEKHKTELESLKHQQDALWTEKLQVLKQQYQTEMEKLR-EKCEQEKETLLKD 659

Query:   401 K------FLQEKDEM-LQELEKKLTQVQNSLLKKEKELEKQQCMATELEMTVKEAKQDKS 453
             K        ++E +E  L++L+ K T+++ SL  +  E+ K +  E E++V + + DK
Sbjct:   660 KEIIFQAHIEEMNEKTLEKLDVKQTELE-SLSSELSEVLKARHKLEE-ELSVLKDQTDKM 717

Query:   454 K-EAECKALQAEVQKLKNSLEEAKQQERLAAQQAAQC-KEEAALAGCHLEDTQRKLQKGL 511
             K E E K +  +    +  ++  ++  ++  Q+  +  K++    L++  +  L++
Sbjct:   718 KQELEAK-MDEQKNHHQQQVDSIIKEHEVSIQRTEKALKDQINQLELLLKERDKHLKEHQ 776

Query:   512 L-LDKQKADTIQELQRELQMLQKESSMAEKEQTSNRKRVEELSLELSEALRKLENSDKEK 570
             ++ +AD I+  + ELQ    +  +  Q++  ++  +  +L++  +KL  +  + E+
Sbjct:   777 AHVENLEAD-IKRSEGELQQASAKLDVFQSYQSATHEQTKAYEEQLAQLQQKLLDLETER 835

Query:   571 RQLQKTVAEQDMKMNDM---LD--RIKHQHREQGSIK--CKLEEDLQEATKLLEDKREQL 623
             L K VAE + +  D+  LD  +I+ Q  Q  K  ++E+ ++  T++ E K E
Sbjct:   836 ILLTKQVAEVEAQKKDVCTELDAHKIQVQDLMQQLEKQNSEMEQKVKSLTQVYESKLEDG 895

Query:   624 KKSKEHEK--LMEGELEALRQEFKKKDKTLKENSRKLEEENENLRAELQCCSTQLESSLN 681
             K +E K L+E E   L+  +K K ++ ++KL + +++  +  T+ ++
Sbjct:   896 NKEQEQTKQILVEKENMILQMREGQK-KEIEILTQKLSAKEDSIHILNEEYETKFKNQEK 954

Query:   682 KYNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAAC 741
             K    +Q  +++ + +    K+ L+  +A+L K L  E  L+ +  ++  ++A  + A
Sbjct:   955 KMEKVKQKAKEMQETL---KKKLLDQEAKLKKEL--ENTALELSQKEKQFNAKMLEMAQA 1009

Query:   742 QD-DLTQALEKLNHVTSETKSLQQSLTQTQEKKAQLEEEIIAYEERMKKLNTELRKLRGF 800
                ++ A+  +L  T++  +  ++ SLT+  +  +L  +  I  +E  KKLN +  +L+
Sbjct:  1010 NSAGISDAVSRLE--TNQKEQIE-SLTEVHRR--ELNDVISIWE---KKLNQQAEELQEI 1061

Query:   801 HQESELEVHAFDKKLEEMSCQVLQW--QKQHQNDLKMLAAKEEQLREFQEEMAALKENLL 858
             H   E+++  ++++ E+  ++L +  +K+  N  ++  KEE +++  +   +  L+E L
Sbjct:  1062 H---EIQLQEKEQEVAELKQKILLFGCEKEEMNK-EITWLKEEGVKQ-DTTLNELQEQLK 1116
```

```
Query:   859 EDDKEPCCLPQWSVPKDTCRLYRGNDQIMTNLEQ--WAKQQKVANEKLGNQLREQVNYI- 915
             +       L Q     K     L +     +  +L++  + ++Q V   + L   + + +V+ +
Sbjct:  1117 QKSAHVNSLAQ-DETKLKAHLEKLEVDLNKSLKENTFLQEQLVELKMLAEEDKRKVSELT 1175

Query:   916 AKLSGEKDHLHSVMVHLQQENKKLK-KEIEEKKMKAE 951
             +KL     +    S+     ++ NK L+ K +E KK+  E
Sbjct:  1176 SKLKTTDEEFQSLKSSHEKSNKSLEDKSLEFKKLSEE 1212

 Score = 337 (50.6 bits), Expect = 4.0e-26, P = 4.0e-26
 Identities = 215/951 (22%), Positives = 433/951 (45%)

Query:    10 REVSSLNSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCMEEAMNSSHDKKQAQALAFEESE 69
             +E  +  +++L L+  ++          K Q K L  +      EA  + H+K+     +  E+ +
Sbjct:   560 QEAETYRTRILELESSLEKSLQENKNQSKDLAVHL----EAEKNKHNKEIT--VMVEKHK 613

Query:    70 VEFGSSKQCHLRQLQQLKKKLLVLQQELEFHTEELQTSYYSLRQYQSILEKQTSDLVLLH 129
              E  S K  H +Q      +KL VL+Q+ +     E+L+            Q +  L K     +++
Sbjct:   614 TELESLK--H-QQDALWTEKLQVLKQQYQTEMEKLREK---CEQEKETLLKD-KEIIFQA 666

Query:   130 HHCKLKE---DEVILYEEEMGNHNENTGEKL---HLAQEQLALAGDKIASLERSLNLYRD 183
             H   ++ E     +++ + + E+ + +       E L      H  +E+L++  D+     +++ L     D
Sbjct:   667 HIEEMNEKTLEKLDVKQTELESLSSELSEVLKARHKLEEELSVLKDQTDKMKQELEAKMD 726

Query:   184 K----YQSSLSNIELLECQVKMLQGE--LGGIMGQEPENKGDHSKVRIYTSPCMIQEHQE 237
             +      +Q  + +I + E +V + + E  L    + Q      + K  +          ++ +
Sbjct:   727 EQKNHHQQQVDSI-IKEHEVSIQRTEKALKDQINQLELLLKERDK-HLKEHQAHVENLEA 784

Query:   238 TQKRLSEVWQKVSQQDDLIQELRNKLACSNALVLEREKALIKLQADFASCTATHRYPPSS 297
              KR     Q+  S + D+  Q   ++        ++          E+ L +LQ              T R
Sbjct:   785 DIKRSEGELQQASAKLDVFQSYQS---ATHEQTKAYEEQLAQLQQKLLDLE-TERIL--- 837

Query:   298 SEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKDMMKL-ELD 356
                   + K +    ++ QK      C    ++  ++    V+DL    +LE   + +        +K + ++ E
Sbjct:   838 ------LTKQVAEVEAQKKDVCTELDAHKIQVQDLMQQLEKQNSEMEQKVKSLTQVYESK 891

Query:   357 LH-GLREETSAHIERKDKDITILQCRL-QELQLEFTETQKLTLKKDKF--LQEKDEM-LQ 411
             L  G +E++             +K+   ILQ R  Q+ ++E   TQKL+ K+D     L E+ E    +
Sbjct:   892 LEDGNKEQEQTKQILVEKENMILQMREGQKKEIEIL-TQKLSAKEDSIHILNEEYETKFK 950

Query:   412 ELEKKLTQVQNSLLK-----KEKELEKQQCMATELEMTVKEAKQDKSKEAECKALQAEVQ 466
              EKK+ +V++    +        K+K L+++  +   ELE T   E  Q K K+  K L+   Q
Sbjct:   951 NQEKKMEKVKQKAKEMQETLKKKLLDQEAKLKKELENTALELSQ-KEKQFNAKMLEM-AQ 1008

Query:   467 KLKNSLEEAKQQERLAAQQAAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKADTIQELQR 526
               + +A     RL   Q  Q +      +     L  D      +K L      Q+A++ +QE+
Sbjct:  1009 ANSAGISDAVS--RLETNQKEQIESLTEVHRRELNDVISIWEKKL---NQQAEELQEIH- 1062

Query:   527 ELQMLQKESSMAEKEQT---------SNRKRV---EELSLELSEALRKLENSDKEKRQLQ 574
             E+Q+ +KE   +AE +Q                 K +      +E  ++        L +L+     K+K
Sbjct:  1063 EIQLQEKEQEVAELKQKILLFGCEKEEMNKEITWLKEEGVKQDTTLNELQEQLKQKSAHV 1122

Query:   575 KTVAEQDMKMNDMLDRIKHQHREQGSIKCKLEEDLQEATKLLEDKREQLKKSKEHEKLME 634
             ++A+ + K+     L++++      +           L+E L E    L E+ + ++ +         K  +
Sbjct:  1123 NSLAQDETKLKAHLEKLEVDLNKSLKENTFLQEQLVELKMLAEEDKRKVSELTSKLKTTD 1182

Query:   635 GELEALRQEFKKKDKTLKENSRKLEEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLN 694
             E ++L+ +    +K +K+L++  S + ++ +E L  +L C  + E+ L    T++ +  +    +
Sbjct:  1183 EEFQSLKSSHEKSNKSLEDKSLEFKKLSEELAIQLDICCKKTEALLEA-KTNELINISSS 1241

Query:   695 KEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDDLT----QALE 750
             K  A+     +   Q + K    KE    ++T     E  +A  R+     Q+ L     QA
Sbjct:  1242 KTNAILSR-ISHCQHRTTKV--KEALLIKTCTVSEL-EAQLRQLTEEQNTLNISFQQATH 1297

Query:   751 KLNHVTSETKSLQQSLTQTQEKKAQLEEEIIAYEERMKKLN---TELRK--LRGFHQESE 805
             +L     ++  KS++   +      +K  L++E        ++      +    T+L+K        +   +
Sbjct:  1298 QLEEKENQIKSMKADIESLVTEKEALQKEGGNQQQAASEKESCITQLKKELSENINAVTL 1357

Query:   806 LEVHAFDKKLE--EMSCQVLQWQKQHQNDLKMLAAKEEQLREFQEEMAALKENLLEDDKE 863
             ++       +KK+E    +S Q+      Q QN +  L+ KE  +     +++      K  LL D +
Sbjct:  1358 MKEELKEKKVEISSLSKQLTDLNVQLQNSIS-LSEKEAAISSLRKQYDEEKCELL-DQVQ 1415

Query:   864 PCCLPQWSVPKDTCRLYRGNDQIMTNLEQWAKQQKVANEKLGNQLRE---QVNYIAKLSG 920
                  ++ K+           D        +W K+ +     +  N ++E    Q+        +K +
Sbjct:  1416 DLSFKVDTLSKEKISALEQVDDWSNKFSEWKKKAQSRFTQHQNTVKELQIQLELKSKEAY 1475

Query:   921 EKDH-LHSVMVHLQQENKK---LKKEIEEKKMKAE 951
             EKD  ++ +   L Q+NK+   LK E+E+ K K E
Sbjct:  1476 EKDEQINLLKEELDQQNKRFDCLKGEMEDDKSKME 1510

 Score = 332 (49.8 bits), Expect = 1.4e-25, P = 1.4e-25
 Identities = 209/953 (21%), Positives = 438/953 (45%)
```

```
Query:    1 MKDEAGERDREVSSLNSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCMEEAMNS----SHD 56
            MK  + E+  ++   L+ K L+ + + +    +  R+R+  +  ++ +E++ +     S +
Sbjct:  470 MKKSSEEQIAKLQKLHEKELARK-EQELTKKLQTREREFQEQMKVALEKSQSEYLKISQE 528

Query:   57 KKQAQALAFEESEVEFGSSKQCHLRQLQQLKKKLLVLQQELEFHTEELQTSYYSLRQYQS 116
            K+Q ++LA EE E++      K+  L + +    KL  LQQE E +    +      SL +
Sbjct:  529 KEQQESLALEELELQ----KKAILTESEN---KLRDLQQEAETYRTRILELESSLEKSLQ 581

Query:  117 ILEKQTSDLVLLHHHCKLKEDE--VILYEE-----EMGNHNENT--GEKLHLAQEQLALA 167
             + Q+ DL +      K K ++   ++ E+      E   H ++     EKL + ++Q
Sbjct:  582 ENKNQSKDLAVHLEAEKNKHNKEITVMVEKHKTELESLKHQQDALWTEKLQVLKQQYQTE 641

Query:  168 GDKIASL--ERSLNLYRDK---YQSSLS--NIELLECQVKMLQGELGGIMGQEPENKGDH 220
            +K+      +    L +DK    +Q+ +   N + LE ++ + Q EL  +  +  E
Sbjct:  642 MEKLREKCEQEKETLLKDKEIIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKAR 700

Query:  221 SKVRIYTSPCMIQEHQETQKRLSEVWQKVSQQDDLIQELRNKLACSNALVLEREKALIKL 280
            K+    S  ++++ +T K   E+  K+ +Q +  Q+  + +   + + ++R +  +K
Sbjct:  701 HKLEEELS--VLKD--QTDKMKQELEAKMDEQKNHHQQQVDSIIKEHEVSIQRTEKALKD 756

Query:  281 QADFASCTATHR--YPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEA 338
            Q +          R  +          E+++  +K  + +        ++ +Q+        + +A
Sbjct:  757 QINQLELLLKERDKHLKEHQAHVENLEADIKRSEGELQQASAKLDVFQSYQSATHEQTKA 816

Query:  339 VSEQKRNIMKDMMKLELDLHGLREETSAHIERKDKDITILQCRLQELQLEFTETQKLTLK 398
              EQ    + + ++ LE +    L ++  A +E + KD+      C    EL       + Q L  +
Sbjct:  817 YEEQLAQLQQKLLDLETERILLTKQV-AEVEAQKKDV----CT--ELDAHKIQVQDLMQQ 869

Query:  399 KDKFLQEKDEMLQELEKKLTQVQNSLLKK-EKELEKQQCMATELEMTVKEAKQDKSKEAE 457
            +K    +   EM Q++ K LTQV  S L++  KE E+ + +  E E  + + ++ + KE E
Sbjct:  870 LEK---QNSEMEQKV-KSLTQVYESKLEDGNKEQEQTKQILVEKENMILQMREGQKKEIE 925

Query:  458 C--KALQAEVQKLKNSLEEAKQQERLAAQQAAQCKEEAAALAGCHLEDTQRK--LQKGLLL 513
             + L A+    +      EE + + +    ++  + K++A         +++T +K  L +    L
Sbjct:  926 ILTQKLSAKEDSIHILNEEYETKFKNQEKKMEKVKQKAK----EMQETLKKKLLDQEAKL 981

Query:  514 DKQKADTIQEL-QRELQMLQKESSMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRQ 572
            K+  +T  EL Q+E Q   K  MA+            V  L      E +  L  ++  +R+
Sbjct:  982 KKELENTALELSQKEKQFNAKMLEMAQANSAGISDAVSRLETNQKEQIESL--TEVHRRE 1039

Query:  573 LQKTVAEQDMKMNDMLDRIKHQHREQGSIKCKLEEDLQEATKLLEDKREQLKKS----KE 628
            L    ++   + K+N    + ++  H  Q   K +  +L++   L   ++E++ K     KE
Sbjct: 1040 LNDVISIWEKKLNQQAEELQEIHEIQLQEKEQEVAELKQKILLFGCEKEEMNKEITWLKE 1099

Query:  629 HEKLMEGELEALRQEFKKKDKTLKENSRKLEEENENLRAELQCCSTQLESSLNKYNTSQQ 688
                 +  L   L+++ K+K   +  NS  L ++   L++A L+       L  SL +    Q+
Sbjct: 1100 EGVKQDTTLNELQEQLKQKSAHV--NS--LAQDETKLKAHLEKLEVDLNKSLKENTFLQE 1155

Query:  689 VIQDLNKEIALQKESLMSLQAQL---DKALQ--KEKHYLQTTITKEA---YDALSRKSAA 740
            + +L        K  +  L ++L  D+  Q  K  H         ++      +  LS +  A
Sbjct: 1156 QLVELKMLAEEDKRKVSELTSKLKTTDEEFQSLKSSHEKSNKSLEDKSLEFKKLSEE-LA 1214

Query:  741 CQDDL----TQAL------EKLNHVTSETKSLQQSLTQTQEKKAQLEEEIIAYEERMKKL 790
            Q D+     T+AL        E +N  +S+T ++   ++  Q +  +++E ++      + +L
Sbjct: 1215 IQLDICCKKTEALLEAKTNELINISSSKTNAILSRISHCQHRTTKVKEALLIKTCTVSEL 1274

Query:  791 NTELRKLRGFHQESELEVHAFDKKLEEMSCQVLQWQKQHQNDLKMLAAKEEQLREFQEEM 850
            +LR+L         +        +LEE   Q+     K  + D++ L  ++E L    Q+E
Sbjct: 1275 EAQLRQLTEEQNTLNISFQQATHQLEEKENQI----KSMKADIESLVTEKEAL---QKEG 1327

Query:  851 AALKENLLEDDKEPCCLPQWSVPKDTCRLYRGNDQIMTNLEQWAKQQKVANEKLGNQLRE 910
              +       +KE C + Q   + K+          N   +T +++   K++KV    L  QL +
Sbjct: 1328 G--NQQQAASEKESC-ITQ--LKKELSE----NINAVTLMKEELKEKKVEISSLSKQLTD 1378

Query:  911 ---QVNYIAKLSGEKDHLHSVMVHLQQENKKLKKEIEEKKMKAE 951
               Q+      LS ++   + S+        +E  +L  ++++     K +
Sbjct: 1379 LNVQLQNSISLSEKEAAISSLRKQYDEEKCELLDQVQDLSFKVD 1422
```

Score = 329 (49.4 bits), Expect = 2.9e-25, P = 2.9e-25
Identities = 226/941 (24%), Positives = 444/941 (47%)

```
Query:   61 QALAFEESEVE--FGSSKQCHLRQLQQLKKKLLVLQQELEFHTEELQTSYYSLRQYQSIL 118
            Q L  E+ +++       S+  LR++ +L+++L + QQ  +   EE  S     QY S+L
Sbjct:  165 QMLQREKKKLQGILSQSQDKSLRRIAELREELQMDQQAKKHLQEEFDASLEEKDQYISVL 224

Query:  119 EKQTSDLVLLHHHCKLKEDEV-----ILYEEEMGNHNENT---GEKL---HLAQEQLALA 167
            + Q S L       + +  D +         +  + E+     EN      GE +      + + L
Sbjct:  225 QTQVSLLKQRLRNGPMNVDVLKPLPQLEPQAEVFTKEENPESDGEPVVEDGTSVKTLETL 284

Query:  168 GDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQGELGGIMGQEPENKGDHSKVRIYT 227
```

```
                 ++   E  L   ++  QS       LL  + + LQ +L   + QE E   D    + +
Sbjct:   285 QQRVKRQENLLKRCKETIQSHKEQCTLLTSEKEALQEQLDERL-QELEKIKD---LHMAE 340

Query:   228 SPCMIQEHQETQKRLSEVWQKVSQQDDLIQELRNKLACSNALVLEREKALIKLQADFASC 287
               +I + ++ +  + ++ Q       +I E + ++      L ++ E+ + +L++
Sbjct:   341 KTKLITQLRDAKNLIEQLEQDKGM---VIAETKRQM--HETLEMKEEE-IAQLRSRIKQM 394

Query:   288 TATH---RYPPSSSEEC--EDIKKILKHLQEQKDSQCLHVEEYQNLVKDL-------RVE 335
             T     R     SE    E+++K L   Q+ ++++       E     +K +        R+
Sbjct:   395 TTQGEELREQEKSERAAFEELEKALSTAQKTEEARRKLKAEMDEQIKTIEKTSEEERIS 454

Query:   336 LEA-VSEQKRNIMKDMMKL--ELDLHGLREETSAHIERKDKDITILQCRLQELQLEFTET 392
             L+  +S  K+ ++ D+MK   E  +  L++       + RK++++T      +LQ  + EF E
Sbjct:   455 LQQELSRVKQEVV-DVMKKSSEEQIAKLQKLHEKELARKEQELTK---KLQTREREFQEQ 510

Query:   393 QKLTLKKDKFLQEKDEMLQELEKKLTQVQNSLLKKEKELEKQQCMATELEMTVKEAKQDK 452
             K+ L+K +    E  ++ QE E+       Q SL  +E EL+K+  + TE E  +++ +Q+
Sbjct:   511 MKVALEKSQ--SEYLKISQEKEQ-----QESLALEELELQKKAIL-TESENKLRDLQQE- 561

Query:   453 SKEAECKALQAEVQKLKNSLEEAKQQER-----LAAQQAAQCKEEAALAGCHLEDTQR-K 506
             ++      + L+ E    L+ SL+E K Q +      L A++      KE    +    H  + + K
Sbjct:   562 AETYRTRILELE-SSLEKSLQENKNQSKDLAVHLEAEKNKHNKEITVMVEKHKTELESLK 620

Query:   507 LQKGLLLDKQKADTIQELQRELQMLQKESSMAEKEQTSNRKRVEELSLELSEALRK-LEN 565
             Q+  L  ++     Q+ Q E++ L +E       EKE     K +         + E   K LE
Sbjct:   621 HQQDALWTEKLQVLKQQYQTEMEKL-REKCEQEKETLLKDKEII-FQAHIEEMNEKTLEK 678

Query:   566 SDKEKRQLQKTVAEQDMKMNDMLDRIKHQHREQGSI-KCKLEEDLQEA-TKLLEDKR--E 621
             D ++ +L++   +E     ++++L + +H+  E+ S+ K + ++  QE   K+ E K   +
Sbjct:   679 LDVKQTELESLSSE----LSEVL-KARHKLEEELSVLKDQTDKMKQELEAKMDEQKNHHQ 733

Query:   622 QLKKS--KEHEKLMEGELEALRQEFKKKDKTLKENSRKLEEEN---ENLRAELQCCSTQL 676
             Q   S  KEHE  ++  +AL+ +  +  + LKE   L+E          ENL A+++   +L
Sbjct:   734 QQVDSIIKEHEVSIQRTEKALKDQINQLELLLKERDKHLKEHQAHVENLEADIKRSEGEL 793

Query:   677 ESSLNKYNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSR 736
             + + K + Q          +++     +E L  LQ +L   L+ E+  L     TK+  +  ++
Sbjct:   794 QQASAKLDVFQSYQSATHEQTKAYEEQLAQLQQKL-LDLETERILL----TKQVAEVEAQ 848

Query:   737 KSAACQD---------DLTQALEKLNHVTSETKSLQQSLTQTQEKKAQ--LEEEIIAYEE 785
             K   C +       DL Q LEK N      SE +   +SLTQ  E K +    +E+     +
Sbjct:   849 KKDVCTELDAHKIQVQDLMQQLEKQN---SEMEQKVKSLTQVYESKLEDGNKEQEQTKQI 905

Query:   786 RMKKLNTELRKLRGFHQESELEVHAFDKKLEEMSCQVL--QWQKQHQNDLKMLAAKEEQL 843
             ++K N  L+    G  Q+ E+E+        +E S  +L  +++ + +N  K +     +++
Sbjct:   906 LVEKENMILQMREG--QKKEIEILTQKLSAKEDSIHILNEEYETKFKNQEKKMEKVKQKA 963

Query:   844 REFQEEMAALKENLLEDDKEPCCLPQWSVPKDTCRLYRGNDQIMTNLEQWAKQQKV---- 899
             +E QE       LK+ LL+ + +     L + +          L +    Q      + + A+
Sbjct:   964 KEMQE---TLKKKLLDQEAK---LKK-ELENTALELSQKEKQFNAKMLEMAQANSAGISD 1016

Query:   900 ANEKLGNQLREQVNYIAKLSG-EKDHLHSVMVH-LQQENKKLKK--EIEEKKMKAENTRL 955
             A  +L    +EQ+  + ++    E + + S+   L Q+ ++L++    EI+ ++ + E    L
Sbjct:  1017 AVSRLETNQKEQIESLTEVHRRELNDVISIWEKKLNQQAEELQEIHEIQLQEKEQEVAEL 1076

Query:   956 CTKALGPSRTESTQREKVCGTLGWKGLPQD 985
               K L       E   +  K     L   +G+ QD
Sbjct:  1077 KQKIL-LFGCEKEEMNKEITWLKEEGVKQD 1105

 Score = 326 (48.9 bits), Expect = 6.0e-25, P = 6.0e-25
 Identities = 220/907 (24%), Positives = 444/907 (48%)

Query:    67 ESEVEFGSSKQCHLRQLQQLKKKLLVLQQELEFHTEELQTSYYSLRQYQSILE---KQTS 123
             E+E  G+S   +   QL Q   +++                EL T+Y   L++ +  L+      Q+
Sbjct:   123 EAEDLVGNSDSLNKEQLIQRLRRMERSLSSYRGKYSELVTAYQMLQREKKKLQGILSQSQ 182

Query:   124 DLVLLHHHCKLKEDEVILYEEEMGNHNENTGEKLHLAQEQLALAGDKIASLERSLNLYRD 183
             D  L      +L+E+  +  +++      H  +    E+   + E+        I+ L+  ++L +
Sbjct:   183 DKSL-RRIAELREE--LQMDQQAKKHLQ---EEFDASLEE---KDQYISVLQTQVSLLKQ 233

Query:   184 KYQSSLSNIELLECQVKMLQGELGGIMGQE-PENKG-----DHSKVR-IYTSPCMIQEHQ 236
             + ++    N+++L+  +  L+  +       +E PE+ G       D + V+ + T      ++ +
Sbjct:   234 RLRNGPMNVDVLK-PLPQLEPQAEVFTKEENPESDGEPVVEDGTSVKTLETLQQRVKRQE 292

Query:   237 ETQKRLSEVWQKVSQQDDLIQELRNKLACSNALVLEREKALIKLQADFASCTATHRYPPS 296
               KR  E Q   +Q  L+        K A     L ER + L K++ D       T
Sbjct:   293 NLLKRCKETIQSHKEQCTLLTS--EKEALQEQLD-ERLQELEKIK-DLHMAEKTKLIT-- 346

Query:   297 SSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKDMMKLELD 356
               +   D K +++ L++ K       +   E    + + + L ++ E ++ Q R+ +K M       +
Sbjct:   347 ---QLRDAKNLIEQLEQDKGM--VIAETKRQMHETLEMKEEEIA-QLRSRIKQMTTQGEE 400
```

```
Query:   357 LHGLREETS-AHIERKDKDITILQCRLQE----LQLEFTETQKLTLKKDKFLQEKDEMLQ 411
             L   +E++  A  E  +K ++  Q + +E    L+ E  E K T++K    +E+  + Q
Sbjct:   401 LREQKEKSERAAFEELEKALSTAQ-KTEEARRKLKAEMDEQIK-TIEKTSE-EERISLQQ 457

Query:   412 ELEKKLTQVQNSLLKK-EKELEKQQCMATELEMTVKEAKQDKSKEAECKALQAEVQKLKN 470
             EL +   +V + + K  E+++ K Q +  E E+  KE  Q+ +K+ + +  + + Q +K
Sbjct:   458 ELSRVKQEVVDVMKKSSEEQIAKLQKLH-EKELARKE--QELTKKLQTREREFQEQ-MKV 513

Query:   471 SLEEAKQQERLAAQQAAQCKEEAALAGCHLEDTQRKLQ-KGLLLD-KQKADTIQELQREL 528
             +LE++ Q E L   Q  + +E   AL    L+      + +  L D +Q+A+T +    EL
Sbjct:   514 ALEKS-QSEYLKISQEKEQQESLALEELELQKKAILTESENKLRDLQQEAETYRTRILEL 572

Query:   529 QMLQKESSMAEKEQTSNRKRVEELSLELSEALRKLENS-DKEKRQLQKTVAEQDMKMNDM 587
             +    E S+ E +  S    V  L  E ++  +++    +K K +L+    +QD    +
Sbjct:   573 ES-SLEKSLQENKNQSKDLAVH-LEAEKNKHNKEITVMVEKHKTELESLKHQQDALWTEK 630

Query:   588 LDRIKHQHR-EQGSIKCKLEEDLQEATKLLEDKRE--QLKKSKEHEKLMEGELEALRQEF 644
             L  +K Q++ E   ++ K E   QE   LL+DK    Q    + +EK +E +L+  + E
Sbjct:   631 LQVLKQQYQTEMEKLREKCE---QEKETLLKDKEIIFQAHIEEMNEKTLE-KLDVKQTEL 686

Query:   645 KKKDKTLKE--NSR-KLEEENENLRAELQCCSTQLESSLNKY-NTSQQVIQDLNKE--IA 698
             +     L E   +R KLEEE   L+ +      +LE+ +++  N  QQ +  + KE  ++
Sbjct:   687 ESLSSELSEVLKARHKLEEELSVLKDQTDKMKQELEAKMDEQKNHHQQQVDSIIKEHEVS 746

Query:   699 LQK-ESLMSLQA-QLDKAL-QKEKHYLQTTITKEAYDALSRKS------AACQDDLTQAL 749
             +Q+ E  +  Q  QL+  L +++KH  +     E  +A  ++S      A+ + D+ Q+
Sbjct:   747 IQRTEKALKDQINQLELLLKERDKHLKEHQAHVENLEADIKRSEGELQQASAKLDVFQSY 806

Query:   750 EKLNHVTSETKSLQQSLTQTQEKKAQLEEEIIAYEERMKKLNTELRKLRGFHQESELEVH 809
             +   H   +TK+ ++ L Q Q+K   LE E I   +++ ++  + + +       +++V
Sbjct:   807 QSATH--EQTKAYEEQLAQLQQKLLDLETERILLTKQVAEVEAQKKDVCTELDAHKIQVQ 864

Query:   810 AFDKKLEEMSCQVLQWQKQHQN--DLKMLAAKEEQLREFQEEMAALKENLL----EDDKE 863
                ++LE+ + ++ Q  K      + K+    +EQ  E  +++   KEN++    E  K+
Sbjct:   865 DLMQQLEKQNSEMEQKVKSLTQVYESKLEDGNKEQ--EQTKQILVEKENMILQMREGQKK 922

Query:   864 PC-CLPQ-WSVPKDTCRLYRGNDQIMTNLE-QWAKQQKVANE--KLGNQLREQV-NYIAK 917
                 L Q  S  +D+  +   N++  T  + Q  K +KV  +  ++   L++++ +  AK
Sbjct:   923 EIEILTQKLSAKEDSIHIL--NEEYETKFKNQEKKMEKVKQKAKEMQETLKKKLLDQEAK 980

Query:   918 LSGEKDHLHSVMVHLQQENKKLKKEIEEKKMKAENTRLCTKALGPSRTESTQREKV 973
             L   K  L +  + L Q+ K+   ++ E  M   N+   + A+  SR E+ Q+E++
Sbjct:   981 L---KKELENTALELSQKEKQFNAKMLE--MAQANSAGISDAV--SRLETNQKEQI 1029

 Score = 318 (47.7 bits), Expect = 4.4e-24, P = 4.4e-24
 Identities = 184/827 (22%), Positives = 405/827 (48%)

Query:     1 MKDEAGERDREVSSLNSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCMEEAMNSSHDKK-Q 59
             ++ E G + +  S   S +  L+ ++    +     ++ L++ ++ +        D   Q
Sbjct:  1323 LQKEGGNQQQAASEKESCITQLKKELSENINAVTLMKEELKEKKVEISSLSKQLTDLNVQ 1382

Query:    60 AQ-ALAFEESEVEFGSSKQCHLRQLQQLKKKLLVLQQELEFHTEELQTSYYS-LRQYQS- 116
              Q +++  E E    S +    +Q  + K +LL   Q+L F  + L     S L Q
Sbjct:  1383 LQNSISLSEKEAAISSLR----KQYDEEKCELLDQVQDLSFKVDTLSKEKISALEQVDDW 1438

Query:   117 ---ILE-KQTSDLVLLHHHCKLKEDEVILYEEEMGNHNENTGEKLHLAQEQLALAGDKIA 172
                  E K+ +      H   +KE ++ L  +    + ++  E+++L +E+L     +
Sbjct:  1439 SNKFSEWKKKAQSRFTQHQNTVKELQIQLELKSKEAYEKD--EQINLLKEELDQQNKRFD 1496

Query:   173 SLERSLNLYRDKYQSSLSNIEL-LECQVKMLQGELGGIMGQEP-ENKGDHSKVRIYTSPC 230
              L+  +   + K +   SN+E  L+ Q   +  EL   + Q+  E +  +  ++ Y
Sbjct:  1497 CLKGEMEDDKSKMEKKESNLETELKSQTARIM-ELEDHITQKTIEIESLNEVLKNYNQQK 1555

Query:   231 MIQEHQETQKRLSEVWQKVSQQDDLIQELRNKLACSNALVLEREKALIKLQADFASCTAT 290
              I EH+E  ++L    +   ++D+ ++E   K+     L LE +   +K + +
Sbjct:  1556 DI-EHKELVQKLQHFQELGEEKDNRVKEAEEKI-----LTLENQVYSMKAELETKKKELE 1609

Query:   291 HRYPPSSSEECEDIKKILKHLQEQKDSQCLHVE-EYQNLVKDLRVELEAVSEQKRNIMKD 349
             H      S+E E++K +   L+ +  ++   ++ + +  +  ++ +L +  E+K    ++
Sbjct:  1610 HVNLSVKSKE-EELKALEDRLESESAAKLAELKRKAEQKIAAIKKQLLSQMEEK----EE 1664

Query:   350 MMKLELDLHGLREETSAHIERKDKDITILQCRLQELQLEFTETQKL--TLKKDKFLQEKD 407
               K   + H    E +  ++ +++++ IL+ +L+ ++   +ET  +  + K      E++
Sbjct:  1665 QYKKGTESH--LSELNTKLQEREREVHILEEKLKSVESSQSETLIVPRSAKNVAAYTEQE 1722

Query:   408 EM-----LQEL-EKKLTQVQNSLLKKEKEL-----EKQQCMATELEMTVK-EAKQDKSKE 455
             E      +Q+  E+K++ +Q +L +KEK L     EK++ +++  EM  + + +  K +
Sbjct:  1723 EADSQGCVQKTYEEKISVLQRNLTEKEKLLQRVGQEKEETVSSHFEMRCQYQERLIKLEH 1782

Query:   456 AECKAL--QAEVQKLKNSLEEAKQQERLAAQQAAQCK--EEAALAGCHLEDTQRKLQKGL 511
```

```
                AE K    Q+ +  L+  LEE  ++  L    Q  + +  +     A  +LE+    +QK L
Sbjct:  1783 AEAKQHEDQSMIGHLQEELEEKNKKYSLIVAQHVEKEGGKNNIQAKQNLENVFDDVQKTL 1842

Query:   512 LLDKQKADTIQELQRELQMLQKESSMAEKEQTSNRKRVEELS--LELSEALRKLENSDKE 569
                 ++K  T Q L+++++ L  +S +  +++  +R  +EEL+   E  +AL++++  +K
Sbjct:  1843 ---QEKELTCQILEQKIKEL--DSCLVRQKEV-HRVEMEELTSKYEKLQALQQMDGRNKP 1896

Query:   570 KRQLQKTVAEQD---MKMNDMLDRIKHQHREQGSIKCKLEEDLQEATKLLEDKREQLKK- 625
                L++    E+     +     +L  ++ QH +         E + Q+  K +    ++ L+
Sbjct:  1897 TELLEENTEEKSKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKEIVRLQKDLRML 1956

Query:   626 SKEHEKLMEGELEALRQEFKKKDKTLKENSRKLEEENENLRAELQCCSTQLESSLNKYNT 685
               KEH++     ELE L++E+ +       E    K+++E E+L  EL+  ST L+  + ++NT
Sbjct:  1957 RKEHQQ----ELEILKKEYDQ------EREEKIKQEQEDL--ELKHNST-LKQLMREFNT 2003

Query:   686 S-QQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDD 744
                Q  Q+L   I          ++A+L ++ Q+E + L   I  E  D L R +A  ++
Sbjct:  2004 QLAQKEQELEMTIKETINKAQEVEAELLESHQEETNQLLKKIA-EKDDDLKR-TAKRYEE 2061

Query:   745 LTQALEKLNHVTSETKSLQQSLTQTQEKKAQ-LEEEIIAYEERMK--KLNTELRKLRGFH 801
               +  A E+   +T++ + LQ  L + Q+K  Q LE+E        + +  +L T+L +
Sbjct:  2062 ILDAREE--EMTAKVRDLQTQLEELQKKYQQKLEQEENPGNDNVTIMELQTQLAQKTTLI 2119

Query:   802 QESELEVHAFDKKLEEMSCQVLQWQK 827
              +S+L++    F +++  +  ++ +++K
Sbjct:  2120 SDSKLKEQEFREQIHNLEDRLKKYEK 2145

 Score = 316 (47.4 bits), Expect = 7.1e-24, P = 7.1e-24
 Identities = 213/977 (21%), Positives = 454/977 (46%)

Query:     4 EAGERD-REVSSLNSKLLSLQLD-IKNLHDVCKRQRKTLQDNQLCMEEAMNSSHDKKQAQ 61
               E   R+  +V S+   K L+ Q + ++ +H++ + Q K  +   +L  +  +    ++   +
Sbjct:  1034 EVHRRELNDVISIWEKKLNQQAEELQEIHEI-QLQEKEQEVAELKQKILLFGCEKEEMNK 1092

Query:    62 ALAFEESEVEFGSSKQCHLRQLQ-QLKKKLL----VLQQE--LEFHTEELQTSYYSLRQY 114
               + + + E   G  +   L +LQ QLK+K      + Q E  L+ H E+L+          +
Sbjct:  1093 EITWLKEE---GVKQDTTLNELQEQLKQKSAHVNSLAQDETKLKAHLEKLEVDLNKSLKE 1149

Query:   115 QSILEKQTSDLVLLHHHCKLKEDEV---ILYEEEMGNHNENTGEKLHLAQEQLALAGDKI 171
               + L++Q  +L +L    K K  E+   +    +E    +++ EK + + E  +L   K+
Sbjct:  1150 NTFLQEQLVELKMLAEEDKRKVSELTSKLKTTDEEFQSLKSSHEKSNKSLEDKSLEFKKL 1209

Query:   172 AS-LERSLNLYRDKYQSSLS--NIELLECQVKMLQGELGGIMGQEPENKGDHSKVRIYTS 228
               +  L   L++   K ++ L       EL+          L  I   +++     K   +
Sbjct:  1210 SEELAIQLDICCKKTEALLEAKTNELINISSSKTNAILSRI--SHCQHRTTKVKEALLIK 1267

Query:   229 PCMIQEHQ-------ETQKRLSEVWQKVSQQ-DDLIQELRNKLACSNALVLEREKALIKL 280
                 C + E +         E Q  L+  +Q+ + Q ++   ++++  A  +LV E+E   L
Sbjct:  1268 TCTVSELEAQLRQLTEEQNTLNISFQQATHQLEEKENQIKSMKADIESLVTEKEA----L 1323

Query:   281 QADFASCTATHRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVS 340
               Q + +          +  S  E C  I ++ K L E  ++  L  EE    +K+ +VE+ ++S
Sbjct:  1324 QKEGGN----QQQAASEKESC--ITQLKKELSENINAVTLMKEE----LKEKKVEISSLS 1373

Query:   341 EQKRNIMKDMMKLELDLHGLREETSAHIERKDKDITILQCRLQEL--QLEFTETQKLT-L 397
               +Q   ++   +     + L        S+  ++ D++  L   ++Q+L  +++    +K++ L
Sbjct:  1374 KQLTDLNVQLQN-SISLSEKEAAISSLRKQYDEEKCELLDQVQDLSFKVDTLSKEKISAL 1432

Query:   398 KK-DKFLQEKDEMLQELEKKLTQVQNSLLKKEKELEKQQCMATELEMTV---KEAKQDKS 453
               ++ D  +  +   E  ++ + + + TQ QN++ + + +LE + A E  + +  KE   ++
Sbjct:  1433 EQVDDWSNKFSEWKKKAQSRFTQHQNTVKELQIQLELKSKEAYEKDEQINLLKEELDQQN 1492

Query:   454 KEAECKALQAEVQKLKNSLEEAKQQERLAAQQAAQCKEEAALAGCHLE-DTQRKLQKGLL 512
               K  +C   +  E  K K  +E+   + L  +Q A   + E   +   +E ++ ++ ++ K
Sbjct:  1493 KRFDCLKGEMEDDKSKMEKKESNLETELKSQTARIMELEDHITQKTIEIESLNEVLKNY- 1551

Query:   513 LDKQKADTIQELQRELQMLQKESSMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRQ 572
                ++QK  +EL  ++LQ   Q+ +     +   +++    L   ++   +LE    KE
Sbjct:  1552 -NQQKDIEHKELVQKLQHFQELGEEKDNRVKEAEEKILTLENQVYSMKAELETKKKELEH 1610

Query:   573 LQKTVAEQDMKMNDMLDRIKHQHREQ-GSIKCKLEEDLQEATKLL----EDKREQLKKSK 627
               +  +V  ++ ++  + DR++ +   +  +K  K E+ +    K L   E+K EQ KK
Sbjct:  1611 VNLSVKSKEEELKALEDRLESESAAKLAELKRKAEQKIAAIKKQLLSQMEEKEEQYKKGT 1670

Query:   628 EHEKLMEGELEALRQEFKKKDKTLKENSRKLEE-ENENL----RAELQCCSTQLESSLNK 682
               E     EL   QE +++  L+E  + +E  ++E L  A+    T+ E + ++
Sbjct:  1671 ESHL---SELNTKLQEREREVHILEEKLKSVESSQSETLIVPRSAKNVAAYTEQEEADSQ 1727

Query:   683 ---YNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSA 739
                    T ++ I  L + +  +KE L++   Q +K  H+  +E  L  A
Sbjct:  1728 GCVQKTYEEKISVLQRNLT-EKEKLLQRVGQ-EKEETVSSHFEMRCQYQERLIKLEHAEA 1785
```

```
Query:   740 ACQDDLTQALEKLNHVTSET--KSLQQSLTQTQEKKAQLEEEIIAYEERMKKLNTELRKL 797
                +D  Q++  + H+  E   K+ + SL   Q  + +  +  I  ++ ++ +  +++K
Sbjct:  1786 KQHED--QSM--IGHLQEELEEKNKKYSLIVAQHVEKEGGKNNIQAKQNLENVFDDVQKT 1841

Query:   798 RGFHQESELEVHAFDKKLEEM-SCQVLQWQKQHQNDLKMLAAKEEQLREFQEEMAALKEN 856
                 QE EL     ++K++E+ SC V Q ++ H+ +++ L +K E+L+  Q+     K
Sbjct:  1842 L---QEKELTCQILEQKIKELDSCLVRQ-KEVHRVEMEELTSKYEKLQALQQMDGRNKPT 1897

Query:   857 -LLEDDKEPCCLPQWSVPKDTCRLYRGNDQIMTNLEQWAKQQKVANEKLGNQLREQVNYI 915
              LLE++ E         PK    +   ++ +   L   A+++K   +KLG ++      +
Sbjct:  1898 ELLEENTEEKSKSHLVQPKLLSNMEAQHNDLEFKLAG-AEREK---QKLGKEIVRLQKDL 1953

Query:   916 AKLSGE-KDHLHSVMVHLQQENK-KLKKEIEEKKMKAENTRLCTKALGPSRTESTQREK 972
               L  E + L  +     QE + K+K+E E+ ++K  +T    + +    T+  Q+E+
Sbjct:  1954 RMLRKEHQQELEILKKEYDQEREEKIKQEQEDLELKHNST--LKQLMREFNTQLAQKEQ 2010

 Score = 301 (45.2 bits), Expect = 2.9e-22, P = 2.9e-22
 Identities = 221/952 (23%), Positives = 441/952 (46%)

Query:     1 MKDEAGERDREVSSLNSKLLSLQLDIKNLHDVCKRQRKTLQDNQL---CMEEAMNSSHD- 56
             +K  A E  R+VS L SKL +   + ++L    ++  K+L+D  L    + E +    D
Sbjct:  1160 LKMLAEEDKRKVSELTSKLKTTDEEFQSLKSSHEKSNKSLEDKSLEFKKLSEELAIQLDI 1219

Query:    57 --KKQAQALAFEESE-VEFGSSK-QCHLRQLQQLKKKLLVLQQELEFHT---EELQTSYY 109
               KK    L  + +E +   SSK    L ++   + +   +++ L   T    EL+
Sbjct:  1220 CCKKTEALLEAKTNELINISSSKTNAILSRISHCQHRTTKVKEALLIKTCTVSELEAQLR 1279

Query:   110 SLRQYQSILEKQTSDLVLLHHHCKLKEDEVILYEEEMGNHNENTGEKLHLAQE---QLAL 166
              L + Q+ L           H  + KE+++   + ++        EK  L +E   Q
Sbjct:  1280 QLTEEQNTLNISFQQAT---HQLEEKENQIKSMKADI---ESLVTEKEALQKEGGNQQQA 1333

Query:   167 AGDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQGELGGIMGQEPENKGDHSKVRIY 226
             A +K    E  +   + +   +++ + L++ ++K  + E+  +  Q  +       V++
Sbjct:  1334 ASEK----ESCITQLKKELSENINAVTLMKEELKEKKVEISSLSKQLTD-----LNVQLQ 1384

Query:   227 TSPCMIQEHQETQKRLSEVWQKVSQQDDLIQELRNKLACSNALVLEREKALIKLQADFAS 286
              S  + ++         +  ++  +  D +Q+L  K+   + L  E+  AL ++  D+++
Sbjct:  1385 NSISLSEKEAAISSLRKQYDEEKCELLDQVQDLSFKV---DTLSKEKISALEQVD-DWSN 1440

Query:   287 CTATHRYPPSS--SEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKD------LRVE-LE 337
               +  +    S  ++    +K++   L E K  +    +E  NL+K+       R + L+
Sbjct:  1441 KFSEWKKKAQSRFTQHQNTVKELQIQL-ELKSKEAYEKDEQINLLKEELDQQNKRFDCLK 1499

Query:   338 AVSEQKRNIM-KDMMKLELDLHGLRE---ETSAHIERKDKDITILQCRLQEL-QLEFTET 392
                E  ++ M K    LE +L        E   HI +K  +I  L   L+   Q +  E
Sbjct:  1500 GEMEDDKSKMEKKESNLETELKSQTARIMELEDHITQKTIEIESLNEVLKNYNQQKDIEH 1559

Query:   393 QKLTLKKDKFLQ---EKDEMLQELEKKLTQVQNSLLKKEKELEKQQCMATELEMTVKEAK 449
             ++L  K   F +   EKD  ++E E+K+  ++N +   + ELE ++     + ++VK
Sbjct:  1560 KELVQKLQHFQELGEEKDNRVKEAEEKILTLENQVYSMKAELETKKKELEHVNLSVK--- 1616

Query:   450 QDKSKEAECKALQAEVQKLKNSLEEAKQQERLAAQQAAQCKEEAALAGCHLEDTQRKLQK 509
                SKE E KAL+  ++    S  +  + +R A Q+ A  K++       +E+ + + +K
Sbjct:  1617 ---SKEEELKALEDRLES--ESAAKLAELKRKAEQKIAAIKKQLL---SQMEEKEEQYKK 1668

Query:   510 GLLLDKQKADT-IQELQRELQMLQKESSMAEKEQTSNRKRVEELSLELSEALRKLENSDK 568
             G      + +T +QE +RE+ +L+++    E  Q+     +   S +   A  + E +D
Sbjct:  1669 GTESHLSELNTKLQEREREVHILEEKLKSVESSQSETL--IVPRSAKNVAAYTEQEEADS 1726

Query:   569 E----KRQLQK-TVAEQDMKMND-MLDRIKHQHREQGSIKCKLEEDLQEATKLLEDKREQ 622
             +    K   +K +V ++++   + +L R+  Q +E+ ++    E   Q   +L+  K E
Sbjct:  1727 QGCVQKTYEEKISVLQRNLTEKEKLLQRVG-QEKEE-TVSSHFEMRCQYQERLI--KLEH 1782

Query:   623 LKKSKEHE-KLMEGEL-EALRQEFKKKDKTLKENSRKLEEENENLRAELQCCSTQLESSL 680
              + +K+HE + M G L E L ++ KK    + ++  K E    N++A+       LE
Sbjct:  1783 AE-AKQHEDQSMIGHLQEELEEKNKKYSLIVAQHVEK-EGGKNNIQAK-----QNLE--- 1832

Query:   681 NKYNTSQQVIQDLNKEIALQKESLMSLQAQLDKAL--QKEKHYLQTTITKEAYDALSR-K 737
             N ++  Q+ +Q+  KE+  Q   L     +LD  L  QKE H ++       Y+ L   +
Sbjct:  1833 NVFDDVQKTLQE--KELTCQ--ILEQKIKELDSCLVRQKEVHRVEMEELTSKYEKLQALQ 1888

Query:   738 SAACQDDLTQALEKLNHVTSETKSLQQSLTQTQEKKAQ-LEEEIIAYEERMKKLNTEL-- 794
                 ++  T+ LE+     S++  +Q  L    E +   LE ++   E   +KL  E+
Sbjct:  1889 QMDGRNKPTELLEENTEEKSKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKEIVR 1948

Query:   795 --RKLRGFHQESELEVHAFDKKLEEMSCQVLQWQKQHQNDLKMLAAKEEQLREFQEEMAA 852
               + LR   +E + E+    K+ ++   + ++ Q+Q   +LK  +  ++ +REF  ++A
Sbjct:  1949 LQKDLRMLRKEHQQELEILKKEYDQEREEKIK-QEQEDLELKHNSTLKQLMREFNTQLAQ 2007

Query:   853 LKENLLEDDKEPCCLPQWSVPKDTCRLYRGNDQIMTNLEQWAKQQKVANEKLGNQLREQV 912
```

```
                 ++ L     KE      Q  V  +     +     Q  TN  Q  K  K+A EK  +  R
Sbjct:  2008 KEQELEMTIKETINKAQ-EVEAELLESH----QEETN--QLLK--KIA-EKDDDLKRTAK 2057

Query:   913 NYIAKLSGEKDHLHSVMVHLQQENKKLKKEIEEKKMKAEN 952
               Y    L    ++ + + +  LQ + ++L+K+ ++K  + EN
Sbjct:  2058 RYEEILDAREEEMTAKVRDLQTQLEELQKKYQQKLEQEEN 2097

 Score = 300 (45.0 bits), Expect = 3.7e-22, P = 3.7e-22
 Identities = 195/961 (20%), Positives = 435/961 (45%)

Query:     1 MKDEAGERDREVSSLNSKLLSLQLDIKN--LHDVCKRQRKTLQDNQLCMEEAMNSSHDKK 58
             +KD+          +  +N K L  +LD+K   L  +      + L+      +EE ++    D+
Sbjct:   657 LKDKEIIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKARHK-LEEELSVLKDQT 714

Query:    59 QAQALAFEESEVEFGSSKQCHLRQLQQLKKKLLV-LQQELEFHTEELQTSYYSLRQYQSI 117
                   +E E +     K  H +Q+  + K+  V +Q+  +    +++         L++
Sbjct:   715 DKMK---QELEAKMDEQKNHHQQVDSIIKEHEVSIQRTEKALKDQINQLELLLKERDKH 771

Query:   118 LEKQTSDLVLLHHHCKLKEDEVILYEEEMG---NHNENTGEKLHLAQEQLALAGDKIASL 174
             L++  + +  L     K  E E+       ++        ++       T E+      +EQLA    K+  L
Sbjct:   772 LKEHQAHVENLEADIKRSEGELQQASAKLDVFQSYQSATHEQTKAYEEQLAQLQQKLLDL 831

Query:   175 ERSLNLYRDKYQSSLSNIELLECQVKMLQGELGGIMGQ-EPENKGDHSKVRIYTSPCMIQ 233
             E      L    +         +  +  +  ++     +  ++       +M Q E +N         KV+  T
Sbjct:   832 ETERILLTKQVAEVEAQKKDVCTELDAHKIQVQDLMQQLEKQNSEMEQKVKSLTQ-VYES 890

Query:   234 EHQETQKRLSEVWQKVSQQDDLIQELRN----KLACSNALVLEREKALIKLQADFASCTA 289
             + ++   K    +  Q + +++++I ++R         ++          +  +E ++  L  ++ +
Sbjct:   891 KLEDGNKEQEQTKQILVEKENMILQMREGQKKEIEILTQKLSAKEDSIHILNEEYET--- 947

Query:   290 THRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKD 349
               ++   +   ++ E +K+  K +QE     + L   E    L K+L      +S++++
Sbjct:   948 --KFK-NQEKKMEKVKQKAKEMQETLKKKLLDQEA--KLKKELENTALELSQKEKQFNAK 1002

Query:   350 MMKL-ELDLHGLREETSA-HIERKDKDITILQCRLQELQLEFTETQKLTLKKDKFLQEKD 407
             M+++ + +  G+ +   S       +K++  ++ +   +EL     + +K   ++ + LQE
Sbjct:  1003 MLEMAQANSAGISDAVSRLETNQKEQIESLTEVHRRELNDVISIWEKKLNQQAEELQEIH 1062

Query:   408 EM-LQELEKKLTQVQNSLLK---KEKELEKQQCMATE----LEMTVKEAKQD-KSKEAEC 458
             E+ LQE E+++ +++    +L      +++E+ K+      E          + T+ E ++  K K A
Sbjct:  1063 EIQLQEKEQEVAELKQKILLFGCEKEEMNKEITWLKEEGVKQDTTLNELQEQLKQKSAHV 1122

Query:   459 KALQAEVQKLKNSLEEAKQQERLAAQQAAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKA 518
              +L    +  KLK   LE+ +       + ++    +E++          E+ +RK+ +   L  K K
Sbjct:  1123 NSLAQDETKLKAHLEKLEVDLNKSLKENTFLQEQLVELKMLAEEDKRKVSE--LTSKLKT 1180

Query:   519 DTIQELQRELQMLQKESSMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRQLQKTVA 578
              T +E Q        +K +      E +       +K   EEL+++L       +K E      + K        + +
Sbjct:  1181 -TDEEFQSLKSSHEKSNKSLEDKSLEFKKLSEELAIQLDICCKKTEALLEAKTN--ELIN 1237

Query:   579 EQDMKMNDMLDRIKH-QHREQGSIKCKLEEDLQEATKLLEDKREQLKKSKEHEKLMEGEL 637
                 K N +L RI H QHR       K++E L   T  + +   QL++ +  E +  +
Sbjct:  1238 ISSSKTNAILSRISHCQHRTT-----KVKEALLIKTCTVSELEAQLRQLTEEQNTLNISF 1292

Query:   638 EALRQEFKKKD---KTLKENSRKLEEENENLR----------AELQCCSTQLESSL---- 680
              +       + ++K+     K++K +     L  E E L+             +E + C TQL+   L
Sbjct:  1293 QQATHQLEEKENQIKSMKADIESLVTEKEALQKEGGNQQQAASEKESCITQLKKELSENI 1352

Query:   681 NKYNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQ-KEKHYLQTTITKEAYDALSRKSA 739
             N       ++ +++     EI+    + L   L  QL   ++      EK        +++ K+ YD       +
Sbjct:  1353 NAVTLMKEELKEKKVEISSLSKQLTDLNVQLQNSISLSEKEAAISSLRKQ-YDEEKCELL 1411

Query:   740 ACQDDLTQALEKLN-HVTSETKSLQQSLTQTQEKKAQLEEEIIAYEERMKKLNTELR-KL 797
                  DL+   ++ L+      S   +  +         +  E  K  + +          ++ +K+L   +L   K
Sbjct:  1412 DQVQDLSFKVDTLSKEKISALEQVDDWSNKFSEWKKKAQSRFTQHQNTVKELQIQLELKS 1471

Query:   798 RGFHQESELEVHAFDKKLEEMSCQVLQWQKQHQNDLKMLAAKEEQLR-EFQEEMAALKEN 856
               +   +++ E +++     ++L++ + +         + + ++D      +   KE     L    E  +  +  A + E
Sbjct:  1472 KEAYEKDE-QINLLKEELDQQNKRFDCLKGEMEDDKSMKEKKESNLETELKSQTARIME- 1529

Query:   857 LLEDDKEPCCLPQWSVPKDTCRLYRGNDQIMTNLEQWAKQQKVANEKLGNQLREQVNYIA 916
              LED              + + T  +    N+ ++ N  Q      QK      K      +L +++ +
Sbjct:  1530 -LEDH----------ITQKTIEIESLNE-VLKNYNQ----QKDIEHK---ELVQKLQHFQ 1570

Query:   917 KLSGEKDH----LHSVMVHLQQENKKLKKEIEEKKMKAENTRLCTKA 959
             +L  EKD+        ++ L+ +    +K E+E KK + E++  L  K+
Sbjct:  1571 ELGEEKDNRVKEAEEKILTLENQVYSMKAELETKKKELEHVNLSVKS 1617

 Score = 298 (44.7 bits), Expect = 6.1e-22, P = 6.1e-22
 Identities = 207/886 (23%), Positives = 412/886 (46%)
```

```
Query:    47 MEEAMNSSHDKKQAQALAFEESEVEFGSSKQCHLRQLQQLKKKLLVLQQELEFHTEELQT 106
             + E  N+ +    Q         EE E +  S K        ++ L   +    LQ+E        +
Sbjct:  1281 LTEEQNTLNISFQQATHQLEEKENQIKSMKA----DIESLVTEKEALQKEGGNQQQAASE 1336

Query:   107 SYYSLRQYQSILEKQTSDLVLLHHHCKLKEDEVILYEEEMGNHNENTGEKLHLAQEQLAL 166
                 + Q +   L +   + + L+     K K+ E+      +++ + N          + L++++ A
Sbjct:  1337 KESCITQLKKELSENINAVTLMKEELKEKKVEISSLSKQLTDLNVQLQNSISLSEKEAA- 1395

Query:   167 AGDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQGELGGIMGQEPENKGDHSKVRIY 226
                  I+SL +       Y ++       L   ++ L  +V  L  E      + Q  +        S+ +
Sbjct:  1396 ----ISSLRKQ---YDEEKCELLDQVQDLSFKVDTLSKEKISALEQVDDWSNKFSEWK-K 1447

Query:   227 TSPCMIQEHQETQKRLS--------EVWQKVSQQDDLIQEL--RNK-LACSNALVLE--- 272
              +       +HQ T K L                E ++K  Q + L +EL   +NK     C        + +
Sbjct:  1448 KAQSRFTQHQNTVKELQIQLELKSKEAYEKDEQINLLKEELDQQNKRFDCLKGEMEDDKS 1507

Query:   273 -REKALIKLQADFASCTAT----HRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQN 327
               EK      L+ +   S TA            +         + E E +  ++LK+   +QKD            E++
Sbjct:  1508 KMEKKESNLETELKSQTARIMELEDHITQKTIEIESLNEVLKNYNQQKDI------EHKE 1561

Query:   328 LVKDLRVELEAVSEQKRNIMKDMMKLELDLHGLREETSAHIERKDKDI--TILQCRLQEL 385
             LV+ L+      + + E+K N +K+    +  L L               A +E K K++      L    + +E
Sbjct:  1562 LVQKLQ-HFQELGEEKDNRVKEAEEKILTLENQVYSMKAELETKKKELEHVNLSVKSKEE 1620

Query:   386 QLEFTETQKLTLKKDKFLQEKDEMLQELEKKLTQVQNSLLKKEKELEKQQCMATELEMTV 445
             +L+     E +       L+ +           +       E+ ++ E+K+    ++     LL + +E E+Q          TE     ++
Sbjct:  1621 ELKALEDR---LESES-AAKLAELKRKAEQKIAAIKKQLLSQMEEKEEQYKKGTESHLSE 1676

Query:   446 KEAKQDKSKEAECKALQAEVQKLKNSLEEAKQQERLAAQQAAQCK-EEAALAGCHLEDTQ 504
                 K     + +E  E       L+ +++ +++ +++S      E         R A     AA    + EEA      GC    +     +
Sbjct:  1677 LNTKLQE-REREVHILEEKLKSVESSQSETLIVPRSAKNVAAYTEQEEADSQGCVQKTYE 1735

Query:   505 RKLQKGLLLDKQKADTIQELQRELQMLQKESSMAEKEQTSNRKRVEELSLELSEALRKLE 564
               K+        +L   +        +     + LQR      Q     +KE      +++           +            R  +   +E     ++L      A      K
Sbjct:  1736 EKIS---VLQRNLTEKEKLLQRVGQ--EKEETVSSHFEM--RCQYQERLIKLEHAEAKQH 1788

Query:   565 NSDKEKRQLQKTVAEQDMKMNDMLDRIKHQHREQG--SIKCK--LE---EDLQ------E 611
                          LQ+ + E++ K + ++                 +H   +E G        +I+ K    LE          +D+Q                E
Sbjct:  1789 EDQSMIGHLQEELEEKNKKYSLIV--AQHVEKEGGKNNIQAKQNLENVFDDVQKTLQEKE 1846

Query:   612 AT-KLLEDKREQELKKSKEHEKLMEG-ELEALRQEFKKKDKTLKENSR----KLEEENENL 665
               T ++LE K ++L              +K +   E+E L   +++K            + + R             +L EEN
Sbjct:  1847 LTCQILEQKIKELDSCLVRQKEVHRVEMEELTSKYEKLQALQQMDGRNKPTELLEENTEE 1906

Query:   666 RAELQCCSTQLESSLN-KYNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQT 724
               +++                 +L   S++     ++N     +     +      +E              + ++ LQ     L  +   L++KE  H   +
Sbjct:  1907 KSKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKEIVRLQKDL-RMLRKE-HQQEL 1964

Query:   725 TITKEAYDALSRKSAACQDDLTQALEKLNHVTSETKSLQQSLTQTQEKKAQLEEEIIAYE 784
               I K+ YD    R+        Q+    + LE L H ++    + +++     TQ      +K+ +LE     I      +
Sbjct:  1965 EILKKEYDQ-EREEKIKQEQ--EDLE-LKHNSTLKQLMREFNTQLAQKEQELEMTI---K 2017

Query:   785 ERMKKLNTELRKLRGFHQESELEVHAFDKKLEEMSCQVLQWQKQHQNDLKMLAAKEEQLR 844
               E + K          +L      HQE       E +     KK+ E          + +     K+++         ++L   A++EE++
Sbjct:  2018 ETINKAQEVEAELLESHQE---ETNQLLKKIAEKDDDLKRTAKRYE---EILDAREEEMT 2071

Query:   845 EFQEEMAALKENLLEDDKEPCCLPQWSVP-KDTCRLYRGNDQIMTNLEQWAKQQKVANEK 903
                   ++       E L +     ++        L Q        P        D         +      ++ T L Q       K        +++ K
Sbjct:  2072 AKVRDLQTQLEELQKKYQQK--LEQEENPGNDNVTIM----ELQTQLAQ--KTTLISDSK 2123

Query:   904 LGNQ-LREQVNYIA-KLSGEKDHLHSVMV-HL 932
               L  Q   REQ++ +   +L      + ++++     V  HL
Sbjct:  2124 LKEQEFREQIHNLEDRLKKYEKNVYATTVGHL 2155

 Score = 280 (42.0 bits), Expect = 5.2e-20, P = 5.2e-20
 Identities = 209/938 (22%), Positives = 432/938 (46%)

Query:     3 DEAGERDREVS-SLNSKLLSLQLDIKN-LHDVC-KRQRKTLQDNQLCMEEAM-NSSHDKK 58
               ++   ++  +E+   +L    KLL         +K   L +        +   +K   Q  N         +E  A    NS+
Sbjct:   957 EKVKQKAKEMQETLKKKLLDQEAKLKKELENTALELSQKEKQFNAKMLEMAQANSAGISD 1016

Query:    59 QAQALAFEESEVEFGSSKQCHLRQLQQLKKKLLVLQQELEFHTEELQTSYYSLRQYQSIL 118
                 L       + E +     S       + H  R+L    +      + + +++L            EELQ        +   ++ +     +
Sbjct:  1017 AVSRLETNQKE-QIESLTEVHRRELNDV---ISIWEKKLNQQAEELQ-EIHEIQLQEK-- 1069

Query:   119 EKQTSDLV--LLHHHCKLKE-DEVILYEEEMGNHNENTGEKLHLAQEQLALAGDKIASLE 175
             E++ ++L         +L        C+ +E ++ I   +  + E G       + T   +L       +Q +       + +A  E
Sbjct:  1070 EQEVAELKQKILLFGCEKEEMNKEITWLKEEGVKQDTTLNELQEQLKQKSAHVNSLAQDE 1129

Query:   176 RSLNLYRDKYQSSLSNIELLECQVKMLQGELGGI--MGQEPENKGDHSKVRIYTSPCMIQ 233
                 L   + +K +    L  N   L  E            LQ +L    +      + +E  + K                ++ T+         Q
```

```
Sbjct: 1130 TKLKAHLEKLEVDL-NKSLKENT--FLQEQLVELKMLAEEDKRKVSELTSKLKTTDEEFQ 1186

Query:  234 E----HQETQKRLSEVWQKVSQQDDLIQELRNKL--AC--SNALVLEREKALIKLQADFA 285
             H+++ K L +    K  +    L +EL  +L    C  + AL+  +    LI + +
Sbjct: 1187 SLKSSHEKSNKSLED---KSLEFKKLSEELAIQLDICCKKTEALLEAKTNELINISSSKT 1243

Query:  286 SCTATH-RYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKR 344
             +   +    +     + + ++    I         + ++Q   + E QN +      +     E+K
Sbjct: 1244 NAILSRISHCQHRTTKVKEALLIKTCTVSELEAQLRQLTEEQNTLNISFQQATHQLEEKE 1303

Query:  345 NIMKDMMKLELD-LHGLREETSAHIERKDKDITILQCRLQELQLEFTET-QKLTLKKDKF 402
             N +K M K +++ L    +E               + +  +  +  +  + +L+ E +E    +TL K++
Sbjct: 1304 NQIKSM-KADIESLVTEKEALQKEGGNQQQAASEKESCITQLKKELSENINAVTLMKEE- 1361

Query:  403 LQEKDEMLQELEKKLTQVQNSLLKKEKELEKQQCMATELEMTVKEAKQDKSKEAECKALQ 462
             L+EK    +  L K+LT + N  L+    L +++    + L      E K +  + +   L
Sbjct: 1362 LKEKKVEISSLSKQLTDL-NVQLQNSISLSEKEAAISSLRKQYDEEKCELLDQVQ--DLS 1418

Query:  463 AEVQKLKNSLEEAKQQERLAAQQAAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKA---- 518
             +V  L        A +Q     + ++ K++A          ++T ++LQ  L L  ++A
Sbjct: 1419 FKVDTLSKEKISALEQVDDWSNKFSEWKKKAQSRFTQHQNTVKELQIQLELKSKEAYEKD 1478

Query:  519 DTIQELQRELQMLQKESSMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRQLQKTVA 578
             + I  L+ EL    K     + E   ++ ++E+    L     +L+++     +L+ +  +
Sbjct: 1479 EQINLLKEELDQQNKRFDCLKGEMEDDKSKMEKKESNLET---ELKSQTARIMELEDHIT 1535

Query:  579 EQDMKMNDMLDRIKHQHREQGSIKCK-LEEDLQEATKLLEDKREQLKKSKEHEKLMEGEL 637
             ++ +++   + + +K+ + +Q  I+ K L + LQ     +L E+K  ++K+++E     +E ++
Sbjct: 1536 QKTIEIESLNEVLKN-YNQQKDIEHKELVQKLQHFQELGEEKDNRVKEAEEKILTLENQV 1594

Query:  638 EALRQEFKKKDKTLKENSRKLEEENENLRAELQCCSTQLES-SLNKYNTSQQVIQDLNKE 696
             +++ E + K K L+  +   ++ + E L+A L+   +LES S  K     ++ +    ++
Sbjct: 1595 YSMKAELETKKKELEHVNLSVKSKEEELKA-LE---DRLESESAAKL---AELKRKAEQK 1647

Query:  697 IALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDDLTQALEKLNHVT 756
             IA  K+ L+S   Q++    +KE+ Y + T  +    L+ K     + ++    EKL  V
Sbjct: 1648 IAAIKKQLLS---QME---EKEEQYKKGT--ESHLSELNTKLQEREREVHILEEKLKSVE 1699

Query:  757 S---ET----KSLQQSLTQTQEKKAQLEEEII-AYEERMKKLNTELRKLRGFHQESELEV 808
             S    ET     +S +     T++++A  +  +   YEE++  L    L         E E  +
Sbjct: 1700 SSQSETLIVPRSAKNVAAYTEQEEADSQGCVQKTYEEKISVLQRNLT-------EKEKLL 1752

Query:  809 HAFDKKLEEMSCQVLQWQKQHQNDLKMLAAKEEQLREFQEEMAALKENLLEDDKEPCCLP 868
             ++ EE      + Q+Q   L    L     E +  E Q  +  L+E L E +K+   +
Sbjct: 1753 QRVGQEKEETVSSHFEMRCQYQERLIKLEHAEAKQHEDQSMIGHLQEELEEKNKKYSLIV 1812

Query:  869 QWSVPKDTCRLYRGNDQIMTNLEQ-WAKQQKVANEK-LGNQLREQ-VNYIAKLSGEKDHL 925
             V K+  +    N Q   NLE  +   QK   EK L  Q+ EQ +  +         +   +
Sbjct: 1813 AQHVEKEGGK---NNIQAKQNLENVFDDVQKTLQEKELTCQILEQKIKELDSCLVRQKEV 1869

Query:  926 HSV-MVHLQQENKKLK 940
             H V M  L  + +KL+
Sbjct: 1870 HRVEMEELTSKYEKLQ 1885

Score = 227 (34.1 bits), Expect = 2.5e-14, P = 2.5e-14
Identities = 160/716 (22%), Positives = 318/716 (44%)

Query:  233 QEHQETQKRLSEVWQKVSQQDDLIQE-LRNKLACSNALV-LEREKALIKL-QADFASCTA 289
             +E  +TQ   ++  +V  + L +   ++  L  S++  L R  +  L + D S TA
Sbjct:   53 RESGDTQSFAQKLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRLDLDSSTA 112

Query:  290 THRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKD 349
             +    P     E ED+       L +++  Q L       +   +   R  +   +   +  + ++
Sbjct:  113 SFDPPSDMDSEAEDLVGNSDSLNKEQLIQRLR--RMERSLSSYRGKYSELVTAYQMLQRE 170

Query:  350 MMKLELDLHGLREETSAHIERKDKDIT-ILQCRLQELQLEFTETQKLTLKKDKFLQEKDE 408
             KL+     G+  +++       +DK +   I + R +ELQ++       + L   + D  L+EKD+
Sbjct:  171 KKKLQ----GILSQS------QDKSLRRIAELR-EELQMDQQAKKHLQEEFDASLEEKDQ 219

Query:  409 MLQELEKKLTQVQNSLLKKEKELEKQQCMATELEMTVKEAKQDKSKEAECKALQAE---V 465
             +  L+ +++ ++   L     ++  + +  +LE    +  ++++ E++  +  +   +   V
Sbjct:  220 YISVLQTQVSLLKQRLRNGPMNVDVLKPLP-QLEPQAEVFTKEENPESDGEPVVEDGTSV 278

Query:  466 QKLKNSLEEAKQQERLA--AQQAAQC-KEEAALAGCHLEDTQRKLQKGLL-LDKQKADTI 521
             + L+    +  K+QE L    ++   Q    KE+   L     E  Q  +L  + L   L+K K     +
Sbjct:  279 KTLETLQQRVKRQENLLKRCKETIQSHKEQCTLLTSEKEALQEQLDERLQELEKIKDLHM 338

Query:  522 QELQRELQMLQKESSMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRQLQKTVAEQD 581
             E  + +  L+   ++  E+ +    + E    ++  E L    E    + R    K +  Q
Sbjct:  339 AEKTKLITQLRDAKNLIEQLEQDKGMVIAETKRQMHETLEMKEEEIAQLRSRIKQMTTQG 398
```

```
Query:   582 MKMNDMLDRIKHQHREQGSIKCKLEEDLQEAT-KLLEDKREQLK---KSKEHEKL-MEGE 636
             ++ +   ++ +      E+           +  +EA  KL   +  EQ+K    K+ E E++ ++ E
Sbjct:   399 EELREQKEKSERAAFEELEKALSTAQKTEEARRKLKAEMDEQIKTIEKTSEEERISLQQE 458

Query:   637 LEALRQEFKK-KDKTLKENSRKLEEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLNK 695
             L  ++QE       K+ +E   KL++ +E     EL      +L   L     T ++  Q+  K
Sbjct:   459 LSRVKQEVVDVMKKSSEEQIAKLQKLHEK---ELARKEQELTKKLQ---TREREFQEQMK 512

Query:   696 EIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDDLTQALEKLN-H 754
              +AL+K      L+    +K  Q+         +  + K+A    S        DL Q  E
Sbjct:   513 -VALEKSQSEYLKISQEKEQQESLALEELELQKKAILTESENKLR---DLQQEAETYRTR 568

Query:   755 VTSETKSLQQSLTQTQEKKAQLEEEIIAYEERMKKLNTELRKLRGFHQESELEV--HAFD 812
               +       SL++SL     QE K Q ++   +   E       K N E+   +    H+ +ELE     H  D
Sbjct:   569 ILELESSLEKSL---QENKNQSKDLAVHLEAEKNKHNKEITVMVEKHK-TELESLKHQQD 624

Query:   813 KKLEEMSCQVLQWQKQHQNDLKMLAAKEEQLRE---------FQEEMAALKENLLED-DK 862
                 E    QVL+   +Q+Q +++ L   K EQ +E            FQ   +  + E   LE  D
Sbjct:   625 ALWTE-KLQVLK--QQYQTEMEKLREKCEQEKETLLKDKEIIFQAHIEEMNEKTLEKLDV 681

Query:   863 EPCCLPQWSVPKDTCRLYRGNDQIMTNLEQWAKQQKVANEKLGNQLREQVNYIAKLSGEK 922
               +    L     S+   +       + +     ++    L         Q         ++L    ++ EQ  N+     +
Sbjct:   682 KQTELE--SLSSELSEVLKARHKLEEELSVLKDQTDKMKQELEAKMDEQKNHHQQQVDSI 739

Query:   923 DHLHSVMVHLQQENKKLKKEIEEKKM 948
                H V +   Q+   K LK +I + ++
Sbjct:   740 IKEHEVSI--QRTEKALKDQINQLEL 763

Score = 183 (27.5 bits), Expect = 1.3e-09, P = 1.3e-09
Identities = 132/584 (22%), Positives = 251/584 (42%)

Query:   409 MLQELEKKLTQVQNSLLKKEKELEKQQCMATELEMTVKEAK-QDKSKEAECKALQAEVQK 467
             M ++L++K+++ Q  L +         +      +T    M  + +    ++   E               + Q
Sbjct:     1 MFKKLKQKISEEQQQLQQALAPAQASSNSSTPTRMRSRTSSFTEQLDEGTPNRESGDTQS 60

Query:   468 LKNSLE-EAKQQERLAAQQAAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKA--DTIQEL 524
             L+       E L       +   ++     +  + R+     L LD     A   D     ++
Sbjct:    61 FAQKLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRLDLDSSTASFDPPSDM 120

Query:   525 QRELQMLQKESSMAEKEQTSNRKRVEELSL-----ELSEALRKLENSDKEKRQLQKTVAE 579
              E + L     S      KEQ   R R  E SL        + SE +    +      +EK++LQ     +++
Sbjct:   121 DSEAEDLVGNSDSLNKEQLIQRLRRMERSLSSYRGKYSELVTAYQMLQREKKKLQGILSQ 180

Query:   580 -QDMKMNDMLDRIKHQHREQGSIKCKLEE---DLQEATK---LLEDKREQLKKSKEHEKL 632
              QD   +   +   +    +      +Q  +  K    EE    L+E   +     +L+  +    LK+     +   +
Sbjct:   181 SQDKSLRRIAELREELQMDQQAKKHLQEEFDASLEEKDQYISVLQTQVSLLKQRLRNGPM 240

Query:   633 MEGELEALRQ-EFKKKDKTLKENSRKLEE---ENENLRAELQCCSTQLESSLNKYNTSQQ 688
               L+  L  Q  E +  +    T  +EN       E      E+          L+    +++        N          ++
Sbjct:   241 NVDVLKPLPQLEPQAEVFTKEENPESDGEPVVEDGTSVKTLETLQQRVKRQENLLKRCKE 300

Query:   689 VIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDDLTQA 748
               IQ     ++    L         +LQ QLD+ LQ  E        ++         E         +++       A      +L +
Sbjct:   301 TIQSHKEQCTLLTSEKEALQEQLDERLQ-ELEKIKDLHMAEKTKLITQLRDA--KNLIEQ 357

Query:   749 LEK-LNHVTSETKSLQQSLTQTQEKKAQLEEEIIAYEERMKKLNTELRKLRGFHQESELE 807
             LE+      V +ETK       + + +T E K       EEEI         R+K++  T++   +LR      Q+ + E
Sbjct:   358 LEQDKGMVIAETK---RQMHETLEMK---EEEIAQLRSRIKQMTTQGEELR--EQKEKSE 409

Query:   808 VHAFDKKLEEMSCQVLQWQKQHQNDLKMLAAKEEQLREFQ----EEMAALKENLLEDDKE 863
                AF      EE+     +     QK      +  K+  A     +EQ++ +      +     EE   +L++ +  L       +E
Sbjct:   410 RAAF----EELEKALSTAQKTEEARRKLKAEMDEQIKTIEKTSEEERISLQQELSRVKQE 465

Query:   864 PCCLPQWSVPKDTCRLYRGNDQIMTNLEQ-WAKQQKVANEKLGNQLR-----EQVNYIAK 917
               + +  S   +     +L + +++ +     EQ        K+ +   +         Q++          Q  Y+ K
Sbjct:   466 VVDVMKKSSEEQIAKLQKLHEKELARKEQELTKKLQTREREFQEQMKVALEKSQSEYL-K 524

Query:   918 LSGEKDHLHSVMVH-LQQENKKLKKEIEEK----KMKAENTRLCTKALGPSRTESTQREK 972
             +S EK+    S+ +   L+ +  K    E E  K        + +AE   R        L   S  +S Q  K
Sbjct:   525 ISQEKEQQESLALEELELQKKAILTESENKLRDLQQEAETYRTRILELESSLEKSLQENK 584
```

Pedant information for DKFZphtes3_1g13, frame 1

Report for DKFZphtes3_1g13.1

```
[LENGTH]      1007
[MW]          117480.77
[pI]          5.90
```

[HOMOL] TREMBL:AF092090_1 product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds. 0.0
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-15
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 5e-15
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 1e-08
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-08
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 1e-08
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1322] 4e-06
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 9e-06
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 5e-04
[EC] 3.6.1.32 Myosin ATPase 1e-16
[PIRKW] nucleus 3e-10
[PIRKW] phosphotransferase 6e-09
[PIRKW] duplication 2e-06
[PIRKW] citrulline 2e-12
[PIRKW] tandem repeat 1e-16
[PIRKW] endocytosis 2e-13
[PIRKW] heart 8e-13
[PIRKW] transmembrane protein 1e-13
[PIRKW] serine/threonine-specific protein kinase 6e-09
[PIRKW] zinc finger 2e-13
[PIRKW] metal binding 2e-13
[PIRKW] DNA binding 4e-12
[PIRKW] muscle contraction 1e-16
[PIRKW] acetylated amino end 1e-11
[PIRKW] actin binding 1e-16
[PIRKW] mitosis 5e-15
[PIRKW] microtubule binding 5e-15
[PIRKW] ATP 1e-16
[PIRKW] thick filament 1e-16
[PIRKW] phosphoprotein 4e-16
[PIRKW] skeletal muscle 2e-14
[PIRKW] calcium binding 2e-12
[PIRKW] alternative splicing 1e-16
[PIRKW] coiled coil 1e-16
[PIRKW] P-loop 1e-16
[PIRKW] heptad repeat 3e-10
[PIRKW] methylated amino acid 1e-16
[PIRKW] immunoglobulin receptor 2e-06
[PIRKW] peripheral membrane protein 2e-13
[PIRKW] cardiac muscle 8e-13
[PIRKW] hydrolase 1e-16
[PIRKW] microtubule 3e-10
[PIRKW] muscle 8e-13
[PIRKW] EF hand 2e-12
[PIRKW] cytoskeleton 2e-15
[PIRKW] hair 2e-12
[PIRKW] calmodulin binding 2e-13
[PIRKW] Golgi apparatus 3e-10
[SUPFAM] myosin heavy chain 1e-16
[SUPFAM] conserved hypothetical P115 protein 1e-07
[SUPFAM] centromere protein E 5e-15
[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 6e-09
[SUPFAM] calmodulin repeat homology 2e-12
[SUPFAM] myosin motor domain homology 1e-16
[SUPFAM] alpha-actinin actin-binding domain homology 2e-07
[SUPFAM] plectin 2e-07
[SUPFAM] trichohyalin 2e-12
[SUPFAM] pleckstrin repeat homology 8e-08
[SUPFAM] ribosomal protein S10 homology 2e-07
[SUPFAM] giantin 3e-13
[SUPFAM] protein kinase homology 6e-09
[SUPFAM] protein kinase C zinc-binding repeat homology 8e-08
[SUPFAM] kinesin motor domain homology 5e-15
[SUPFAM] human early endosome antigen 1 2e-13
[SUPFAM] M5 protein 1e-07
[PROSITE] LEUCINE_ZIPPER 7
[PROSITE] MYRISTYL 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 20

```
[PROSITE]       TYR_PHOSPHO_SITE        1
[PROSITE]       PKC_PHOSPHO_SITE        16
[PROSITE]       ASN_GLYCOSYLATION       2
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY      15.00 %
[KW]            COILED_COIL         42.40 %

SEQ     MKDEAGERDREVSSLNSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCMEEAMNSSHDKKQA
SEG     ............xxxxxxxxxxxx....................................
PRD     ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   .............CCCCCCCCCCCCCCCCCCCCCCCCCCCCC..................

SEQ     QALAFEESEVEFGSSKQCHLRQLQQLKKKLLVLQQELEFHTEELQTSYYSLRQYQSILEK
SEG     ...............xxxxxxxxxxxxxxxxxxxxxx.......................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC............

SEQ     QTSDLVLLHHHCKLKEDEVILYEEEMGNHNENTGEKLHLAQEQLALAGDKIASLERSLNL
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   .............................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ     YRDKYQSSLSNIELLECQVKMLQGELGGIMGQEPENKGDHSKVRIYTSPCMIQEHQETQK
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCCCCCCCCCCCCC.........................................

SEQ     RLSEVWQKVSQQDDLIQELRNKLACSNALVLEREKALIKLQADFASCTATHRYPPSSSEE
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ............................................................

SEQ     CEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKDMMKLELDLHGL
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   .........CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ     REETSAHIERKDKDITILQCRLQELQLEFTETQKLTLKKDKFLQEKDEMLQELEKKLTQV
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCC...........................................CCCCCCCCCCCCCC

SEQ     QNSLLKKEKELEKQQCMATELEMTVKEAKQDKSKEAECKALQAEVQKLKNSLEEAKQQER
SEG     ...xxxxxxxxxx.......................................xxxxxxxx
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ     LAAQQAAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKADTIQELQRELQMLQKESSMAEK
SEG     xxxxxxxxxxxxxxxx.......xxxxxxxxxxxxxx.......................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCC....................................................

SEQ     EQTSNRKRVEELSLELSEALRKLENSDKEKRQLQKTVAEQDMKMNDMLDRIKHQHREQGS
SEG     .........xxxxxxxxxxxxxxx....................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....................

SEQ     IKCKLEEDLQEATKLLEDKREQLKKSKEHEKLMEGELEALRQEFKKKDKTLKENSRKLEE
SEG     ..................................................xxxxxxxxxx
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   .................................................CCCCCCCCCCC

SEQ     ENENLRAELQCCSTQLESSLNKYNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKH
SEG     xxxxxxxxx...................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCCCCCCCCCCCC.....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ     YLQTTITKEAYDALSRKSAACQDDLTQALEKLNHVTSETKSLQQSLTQTQEKKAQLEEEI
SEG     ...................................xxxxxxxxxxxxxxxxxx.......
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   .....................................CCCCCCCCCCCCCCCCCCCCCCC

SEQ     IAYEERMKKLNTELRKLRGFHQESELEVHAFDKKLEEMSCQVLQWQKQHQNDLKMLAAKE
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCCCCCCCCCCCCCCCCC.....................................

SEQ     EQLREFQEEMAALKENLLEDDKEPCCLPQWSVPKDTCRLYRGNDQIMTNLEQWAKQQKVA
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
```

```
COILS  ..........................................................

SEQ    NEKLGNQLREQVNYIAKLSGEKDHLHSVMVHLQQENKKLKKEIEEKKMKAENTRLCTKAL
SEG    ..................................xxxxxxxxxxxxxxxxx.........
PRD    hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS  ............CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC................

SEQ    GPSRTESTQREKVCGTLGWKGLPQDMGQRMDLTKYIGMPHCPGSSYC
SEG    ...............................................
PRD    cchhhhhhhhhhhhhhhhcccccccccchhhhhheeecccccccccc
COILS  ...............................................
```

Prosite for DKFZphtes3_1g13.1

PS00001	52->56	ASN_GLYCOSYLATION	PDOC00001
PS00001	684->688	ASN_GLYCOSYLATION	PDOC00001
PS00004	240->244	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	415->419	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	74->77	PKC_PHOSPHO_SITE	PDOC00005
PS00005	110->113	PKC_PHOSPHO_SITE	PDOC00005
PS00005	238->241	PKC_PHOSPHO_SITE	PDOC00005
PS00005	290->293	PKC_PHOSPHO_SITE	PDOC00005
PS00005	392->395	PKC_PHOSPHO_SITE	PDOC00005
PS00005	396->399	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00005	503->506	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	566->569	PKC_PHOSPHO_SITE	PDOC00005
PS00005	600->603	PKC_PHOSPHO_SITE	PDOC00005
PS00005	650->653	PKC_PHOSPHO_SITE	PDOC00005
PS00005	655->658	PKC_PHOSPHO_SITE	PDOC00005
PS00005	735->738	PKC_PHOSPHO_SITE	PDOC00005
PS00005	876->879	PKC_PHOSPHO_SITE	PDOC00005
PS00005	968->971	PKC_PHOSPHO_SITE	PDOC00005
PS00006	39->43	CK2_PHOSPHO_SITE	PDOC00006
PS00006	53->57	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	296->300	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	444->448	CK2_PHOSPHO_SITE	PDOC00006
PS00006	471->475	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	536->540	CK2_PHOSPHO_SITE	PDOC00006
PS00006	566->570	CK2_PHOSPHO_SITE	PDOC00006
PS00006	576->580	CK2_PHOSPHO_SITE	PDOC00006
PS00006	650->654	CK2_PHOSPHO_SITE	PDOC00006
PS00006	674->678	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	888->892	CK2_PHOSPHO_SITE	PDOC00006
PS00006	963->967	CK2_PHOSPHO_SITE	PDOC00006
PS00006	968->972	CK2_PHOSPHO_SITE	PDOC00006
PS00007	135->143	TYR_PHOSPHO_SITE	PDOC00007
PS00008	207->213	MYRISTYL	PDOC00008
PS00008	599->605	MYRISTYL	PDOC00008
PS00029	83->105	LEUCINE_ZIPPER	PDOC00029
PS00029	90->112	LEUCINE_ZIPPER	PDOC00029
PS00029	97->119	LEUCINE_ZIPPER	PDOC00029
PS00029	104->126	LEUCINE_ZIPPER	PDOC00029
PS00029	403->425	LEUCINE_ZIPPER	PDOC00029
PS00029	410->432	LEUCINE_ZIPPER	PDOC00029
PS00029	918->940	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_1g13.1)

DKFZphtes3_1k11
----

group: cell structure and motility

DKFZphtes3_1k11 encodes a novel 589 amino acid protein with strong similarity to Mus musculus actin-binding protein (ENC-1).

Ectoderm-neural cortex-1 protein (ENC-1) is an early and highly specific marker of neural induction in vertebrates. The protein is related to the kelch family proteins and is expressed during early gastrulation in the prospective neuroectodermal region of the epiblast and later in development throughout the nervous system (NS). ENC-1 functions as an actin-binding protein organising the actin cytoskeleton during neural differentiation and development of the NS. The novel protein is highly similar to ENC-1.

The new protein can find application in modulation of cyto skeleton organisation in human testicular cells.

strong similarity to mouse ENC-1

complete cDNA, compete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3525 bp
Poly A stretch at pos. 3515, polyadenylation signal at pos. 3499

```
   1 GGTGGAGAGC CGGCCGACGG GAGCCGCGGC GGAGCCTGTT GAGCTCGCGC
  51 GGGCTGCCGG GAGTGGTCTC TGAGGCGGCG GCGGCGGCGG GGATCGTCTC
 101 CGGCACTGGC GCACCATGTC GGTCAGTGTC CATGAGACCC GCAAGTCGCG
 151 GAGCAGCACG GGGTCCATGA ACGTCACCCT CTTCCACAAG GCCTCCCACC
 201 CGGACTGTGT GCTGGCCCAC CTCAACACGC TTCGCAAGCA CTGCATGTTC
 251 ACCGACGTCA CACTCTGGGC GGGCGACCGT GCCTTCCCCT GTCACCGTGC
 301 CGTGCTGGCC GCCTCTAGCC GCTATTTTGA GGCCATGTTC AGCCATGGCC
 351 TTCGGGAGAG CCGGGATGAC ACTGTCAACT TCCAGGACAA CCTGCACCCG
 401 GAGGTGCTGG AGCTGCTGCT GGACTTTGCC TACTCCTCAC GCATCGCCAT
 451 CAACGAGGAG AACGCTGAGT CACTGCTGGA GGCAGGCGAC ATGCTGCAGT
 501 TCCACGATGT GCGGGATGCT GCCGCCGAGT TCCTGGAGAA GAACCTTTTC
 551 CCCTCCAACT GCCTGGGCAT GATGCTGCTC TCGGACGCCC ACCAGTGCCG
 601 CCGGCTGTAT GAGTTCTCCT GGCGCATGTG CCTGGTGCAC TTTGAGACGG
 651 TGAGGCAGAG CGAGGACTTC AACAGCCTGT CCAAGGACAC ACTGCTGGAC
 701 CTCATCTCGA GTGATGAGCT GGAGACCGAG GACGAGCGGG TGGTCTTCGA
 751 GGCCATCCTC CAGTGGGTGA AGCACGACCT GGAGCCACGG AAGGTCCACT
 801 TGCCCGAGCT CCTCCGCAGC GTGCGTCTGG CCTTGCTGCC GTCCGACTGC
 851 CTGCAGGAGG CCGTCTCCAG CGAGGCCCTC CTCATGGCAG ACGAGCGCAC
 901 CAAGCTTATC ATGGATGAGG CCCTGCGCTG CAAGACCAGG ATCCTGCAGA
 951 ATGATGGCGT GGTCACCAGC CCCTGTGCCC GGCCACGCAA GGCGGGCCAC
1001 ACGCTACTCA TCCTGGGGGG CCAGACCTTC ATGTGTGACA AGATCTACCA
1051 GGTGGACCAC AAGGCCAAGG AGATCATCCC CAAGGCCGAC CTGCCCAGCC
1101 CCCGGAAGGA GTTCAGCGCC TCAGCGATCG GCTGCAAGGT CTATGTGACG
1151 GGGGGCAGGG GCTCCGAGAA CGGGGTCTCC AAGGATGTCT GGGTGTACGA
1201 CACCGTACAT GAGGAATGGT CCAAGGCGGC GCCCATGCTG ATTGCCCGCT
1251 TTGGCCATGG CTCAGCTGAG CTGGAGAACT GCCTCTATGT GGTGGGGGGA
1301 CACACATCCC TGGCAGGGGT CTTCCCGGCC TCGCCTTCTG TCTCCCTGAA
1351 ACAAGTGGAG AAATACGACC CTGGGGCCAA CAAGTGGATG ATGGTGGCCC
1401 CCTTGCGGGA TGGCGTCAGC AATGCCGCAG TGGTGAGTGC CAAGCTGAAG
1451 CTCTTTGTTT TCGGAGGAAC CAGCATCCAC CGGGACATGG TGTCCAAGGT
1501 CCAGTGCTAT GACCCCTCGG AGAACAGGTG GACGATCAAG GCCGAGTGCC
1551 CCCAGCCTTG GCGGTACACA GCCGCTGCCG TCCTGGGCAG CCAGATCTTC
1601 ATCATGGGAG GTGACACGGA ATTCACAGCC GCCTCGGCCT ACCGCTTTGA
1651 CTGTGAGACC AACCAGTGGA CGCGGATTGG GGACATGACT GCCAAGCGCA
1701 TGTCCTGCCA TGCCCTGGCT TCCGGCAACA AGCTCTATGT GGTCGGGGGC
1751 TACTTTGGGA CCCAGAGGTG TAAGACTCTG GACTGCTATG ACCCCACTTC
1801 AGATACATGG AACTGCATCA CCACAGTGCC CTACTCACTT ATCCCCACGG
1851 CCTTTGTCAG CACCTGGAAG CACCTGCCCG CGTGAGGAGC ACCTGCTGAG
1901 CCCAGCCAGA CCGCGGCCTT CAGTGTCACA GCGTGGCCTT GCTTGTCTGC
1951 CACAGCGGGA GCTAAGCCGG CCCTGGGCCA GCACTCCGAG AGGTGGAAGG
2001 GGCCCTGCCA GCTCTGGGGA GCAGCAGCCT TGGGCTGTTC TGAGCTTTAG
2051 GCAAGAGAAG AGAAGCATCT CTTGCATCCG TGCCCCTGGG GGCCTCTTCA
2101 GCTTTGCAGT GGTTTGTGGG AAGACATACC TCCCAGAGGG GCATGGACTG
2151 CCACCAGGAC TGACCCTGGC GTCGGGGAGA AGGACACTTG CAGAGCCTTG
2201 AGATCACCTG TTTGGCAGGT CCTGGACTGG GGCCGGGCAG GCAGGGGCAG
2251 GGAGGCGCCC CGGGTGGGCT TTGGGGCTGC GGCACTGCCA CACATCCTTT
2301 CCCTCCTGGC CTGCCCTGCT GGGGCTCTAC TGCCATCTAT AGATGGTGTC
2351 CTGGGCCTGG GAAACTAGGT TCCCAGGGGT TGAGACCAGA AAGGTGACCA
2401 AGACAGATTT TTTAAGGTGC AGAAACTGCA GGGGGGCCTC AGTGACATCC
2451 ATGAGGCCTT ATTAGCAAAG GACACCCAGA CCTCCAAGGT TTGTGGGCCC
2501 CTTCCACAAA GCTGTAAGTC CCAGCCCACC TACTCAGGGC CTTGCTCAGT
2551 GCTGTGGCCC GGTGGGGACA CAGTTGCTCG TGGCCACTCA GTGGAGCTGG
2601 GCCTGCAGCA GACTCAAGGC TCCGAGTGCC CTGGGGGTCA CCCCTCCCCT
2651 CCCCTCCTCA GAGCCCACCC TGAGAGGCAG CAGTGACCCC CATGGCACAC
2701 ACCTGCCAAC AGCACTGGGG GCTTCTCCCC AGGAGACCAC GCTGCCCTCC
```

```
2751 AAGACCAGGA GCAGCTGTGA GCTGGAGACA GCAGAGGGAC CCCAGGGTGT
2801 CCCCTGCAGA TCCCACCAGG GCCGCATCCA TCTCAGTGTG GAGGACAGTG
2851 ACGGGACCCT CACCATCCTC TTGCGTTTTG GCCCCCATTT GCTCCCTGAG
2901 CTCCAAGATA AGAATGGCCC CGAGAGAACT GCTGAACATT TGTTCATTGC
2951 TGTCACCTCC TGAGTCACTG GGGTCCCTCA CCAGCACCTC CCTGACACCT
3001 GGGCTATGGA GAGGTTGGCG CCTGTCAGTG ACCATCCTAA TGCCTCTCGC
3051 TCACTCCCAA GCCACCATTT GAGAGGGAGG GGTGTTGGTG CCCTGACAGG
3101 GACTGGGCAG GGTGTCCAAA CTTGGGGCTT CCCAGGCACC TGCAGTGTGA
3151 ACACTGCTTG GCTGGCTCAA GATTAGGGCC GCGGAGGGGG CTGTGCACAT
3201 ACCAGTTACT TAAGCAGCCA CGAGTGTCCC CCATGCCTTG GTGCGGGTCC
3251 TGGAGGCCTC TTGGGGGTGG GACCTTTGGG CAGGGTTTGC CCACTGACGC
3301 GCCCGCCATG GGGCACTGGC TGCATGGGGC TCCTTGGACC CTGTAGACCC
3351 AGCAGGAGCC TGGCCGCGGG GACTGCAGGG AGGGTGCCTG GACCCGTGGG
3401 GTTGCTTCAT TGAGATAAAG CACACTTATC ACATAGCACA AAGGACGTGC
3451 CATGGTGCTT TCCCCAAAAG TTGTGTTGCT TTTATCAGTT TTCTAACTTA
3501 ATAAAAAGAG TTGAGAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

98350113:
Cloning of human ENC-1 and evaluation of its expression and regulation in nervous system tumors.

97252647:
ENC-1: a novel mammalian kelch-related gene specifically expressed in the nervous system
encodes an actin-binding protein.

98234394:
NRP/B, a novel nuclear matrix protein, associates with p110(RB) and is involved in neuronal differentiati

Peptide information for frame 2

ORF from 116 bp to 1882 bp; peptide length: 589
Category: strong similarity to known protein
Classification: Cell structure/motility

```
  1 MSVSVHETRK SRSSTGSMNV TLFHKASHPD CVLAHLNTLR KHCMFTDVTL
 51 WAGDRAFPCH RAVLAASSRY FEAMFSHGLR ESRDDTVNFQ DNLHPEVLEL
101 LLDFAYSSRI AINEENAESL LEAGDMLQFH DVRDAAAEFL EKNLFPSNCL
151 GMMLLSDAHQ CRRLYEFSWR MCLVHFETVR QSEDFNSLSK DTLLDLISSD
201 ELETEDERVV FEAILQWVKH DLEPRKVHLP ELLRSVRLAL LPSDCLQEAV
251 SSEALLMADE RTKLIMDEAL RCKTRILQND GVVTSPCARP RKAGHTLLIL
301 GGQTFMCDKI YQVDHKAKEI IPKADLPSPR KEFSASAIGC KVYVTGGRGS
351 ENGVSKDVWV YDTVHEEWSK AAPMLIARFG HGSAELENCL YVVGGHTSLA
401 GVFPASPSVS LKQVEKYDPG ANKWMMVAPL RDGVSNAAVV SAKLKLFVFG
451 GTSIHRDMVS KVQCYDPSEN RWTIKAECPQ PWRYTAAAVL GSQIFIMGGD
501 TEFTAASAYR FDCETNQWTR IGDMTAKRMS CHALASGNKL YVVGGYFGTQ
551 RCKTLDCYDP TSDTWNCITT VPYSLIPTAF VSTWKHLPA
```

BLASTP hits

Entry MMU65079_1 from database TREMBL:
gene: "ENC-1"; product: "actin-binding protein"; Mus musculus actin-binding protein (ENC-1) mRNA, complete cds.
Score = 2402, P = 1.9e-249, identities = 440/589, positives = 513/589

Entry AF059611_1 from database TREMBLNEW:
gene: "NRPB"; product: "nuclear matrix protein NRP/B"; Homo sapiens nuclear matrix protein NRP/B (NRPB) mRNA, complete cds.
Score = 2400, P = 3.0e-249, identities = 440/589, positives = 512/589

Entry AF010314_1 from database TREMBL:
gene: "PIG10"; product: "Pig10"; Homo sapiens Pig10 (PIG10) mRNA, complete cds.
Score = 1745, P = 7.8e-180, identities = 335/507, positives = 403/507

Entry KELC_DROME from database SWISSPROT:
RING CANAL PROTEIN (KELCH PROTEIN). >TREMBL:DMRCPA_1 product: "ring canal protein"; Drosophila melanogaster ring canel protein and ORF2 mRNA, complete cds.
Score = 672, P = 3.9e-66, identities = 168/536, positives = 257/536

Alert BLASTP hits for DKFZphtes3_1k11, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_1k11, frame 2

Report for DKFZphtes3_1k11.2

```
[LENGTH]        589
[MW]            65923.45
[pI]            6.10
[HOMOL]         TREMBL:MMU65079_1 gene: "ENC-1"; product: "actin-binding protein"; Mus musculus
actin-binding protein (ENC-1) mRNA, complete cds. 0.0
[FUNCAT]        10.05.99 other pheromone response activities       [S. cerevisiae, YHR158c]
2e-09
[BLOCKS]        BL01016D Glycoprotease family proteins
[PIRKW]         zinc finger 1e-08
[PIRKW]         DNA binding 1e-08
[PIRKW]         transcription factor 1e-08
[SUPFAM]        POZ domain homology 3e-68
[SUPFAM]        vaccinia virus 59K HindIII-C protein 1e-15
[SUPFAM]        A55R protein 5e-29
[SUPFAM]        hypothetical protein YHR158c 4e-08
[SUPFAM]        A55R protein middle region homology 5e-29
[SUPFAM]        myxoma virus M9-R protein 1e-14
[SUPFAM]        A55R protein carboxyl-terminal homology 5e-29
[KW]            Alpha_Beta

SEQ     MSVSVHETRKSRSSTGSMNVTLFHKASHPDCVLAHLNTLRKHCMFTDVTLWAGDRAFPCH
PRD     cccccccccccccccccceeeeeeccccchhhhhhhhhhhhhhhheeeeeeccccchhhh

SEQ     RAVLAASSRYFEAMFSHGLRESRDDTVNFQDNLHPEVLELLLDFAYSSRIAINEENAESL
PRD     hccccccccccccccccchhhhhheeeeccccchhhhhhhhhhhhccceeehhhhhhhh

SEQ     LEAGDMLQFHDVRDAAAEFLEKNLFPSNCLGMMLLSDAHQCRRLYEFSWRMCLVHFETVR
PRD     hhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ     QSEDFNSLSKDTLLDLISSDELETEDERVVFEAILQWVKHDLEPRKVHLPELLRSVRLAL
PRD     hhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhc

SEQ     LPSDCLQEAVSSEALLMADERTKLIMDEALRCKTRILQNDGVVTSPCARPRKAGHTLLIL
PRD     ccchhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhcccccccccccccccccceeeeee

SEQ     GGQTFMCDKIYQVDHKAKEIIPKADLPSPRKEFSASAIGCKVYVTGGRGSENGVSKDVWV
PRD     ccccccccceeeecccccccccccccccccccceeeeeeceeeeeecccccccccceeeee

SEQ     YDTVHEEWSKAAPMLIARFGHGSAELENCLYVVGGHTSLAGVFPASPSVSLKQVEKYDPG
PRD     ccccccccccccccccccccceeecccceeeeecccccccccccccccccccceeecccc

SEQ     ANKWMMVAPLRDGVSNAAVVSAKLKLFVFGGTSIHRDMVSKVQCYDPSENRWTIKAECPQ
PRD     ccceeeecccccccceeeeecccceeeeeccccccccccceeeecccccccccccccccc

SEQ     PWRYTAAAVLGSQIFIMGGDTEFTAASAYRFDCETNQWTRIGDMTAKRMSCHALASGNKL
PRD     cccccceeeeecceeeeeccccccccccceeeccccccccceeecccccccccceeeeeecccee

SEQ     YVVGGYFGTQRCKTLDCYDPTSDTWNCITTVPYSLIPTAFVSTWKHLPA
PRD     eeeccccccccccccccccccccceeeeeccccccceeeeeeeccccc
```

(No Prosite data available for DKFZphtes3_1k11.2)

(No Pfam data available for DKFZphtes3_1k11.2)

DKFZphtes3_1n3
----

group: signal transduction

DKFZphtes3_1n3 encodes a novel 1196 amino acid protein with similarity to S. pombe Tupl protein.

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a RGD site is present.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to Tup1p

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="6q24"

Insert length: 5277 bp
Poly A stretch at pos. 5267, polyadenylation signal at pos. 5244

```
   1 GCTGCATAAA GCTGAGAGAT GCCTACAGCT GAGAGTGAAG CAAAAGTAAA
  51 AACCAAAGTT CGCTTTGAAA AATTGCTTAA GACCCACAGT GATCTAATGC
 101 GTGAAAAGAA AAAACTGAAG AAAAAACTTG TCAGGTCTGA AGAAAACATC
 151 TCACCTGACA CTATTAGAAG CAATCTTCAC TATATGAAAG AAACTACAAG
 201 TGATGATCCC GACACTATTA GAAGCAATCT TCCCCATATT AAAGAAACTA
 251 CAAGTGATGA TGTAAGTGCT GCTAACACTA ACAACCTGAA GAAGAGCACG
 301 AGAGTCACTA AAAACAAATT GAGGAACACA CAGTTAGCAA CTGAAAATCC
 351 TAATGGTGAT GCTAGTGTAG AGGAAGACAA ACAAGGAAAG CCAAATAAAA
 401 AGGTGATAAA GACGGTGCCC CAGTTGACTA CACAAGACCT GAAACCGGAA
 451 ACTCCTGAGA ATAAGGTTGA TTCTACACAC CAGAAAACAC ATACAAAGCC
 501 ACAGCCAGGC GTTGATCATC AGAAAAGTGA GAAGGCAAAT GAGGGAAGAG
 551 AAGAGACTGA TTTAGAAGAG GATGAAGAAT TGATGCAAGC ATATCAGTGC
 601 CATGTAACTG AAGAAATGGC AAAGGAGATT AAGAGGAAAA TAAGAAAGAA
 651 ACTGAAAGAA CAGTTGACTT ACTTTCCCTC AGATACTTTA TTCCATGATG
 701 ACAAACTAAG CAGTGAAAAA AGGAAAAAGA AAAAGGAAGT TCCAGTCTTC
 751 TCTAAAGCTG AAACAAGTAC ATTGACCATC TCTGGTGACA CAGTTGAAGG
 801 TGAACAAAAG AAAGAATCTT CAGTTAGATC AGTTTCTTCA GATTCTCATC
 851 AAGATGATGA AATAAGCTCA ATGGAACAAA GCACAGAAGA CAGCATGCAA
 901 GATGATACAA AACCTAAACC AAAAAAAACA AAAAAGAAGA CTAAAGCAGT
 951 TGCAGATAAT AATGAAGATG TTGATGGTGA TGGTGTTCAT GAAATAACAA
1001 GCCGAGATAG CCCGGTTTAT CCCAAATGTT TGCTTGATGA TGACCTTGTC
1051 TTGGGAGTTT ACATTCACCG AACTGATAGA CTTAAGTCAG ATTTTATGAT
1101 TTCTCACCCA ATGGTAAAAA TTCATGTGGT TGATGAGCAT ACTGGTCAAT
1151 ATGTCAAGAA AGATGATAGT GGACGGCCTG TTTCATCTTA CTATGAAAAA
1201 GAGAATGTGG ATTATATTCT TCCTATTATG ACCCAGCCAT ATGATTTTAA
1251 ACAGTTAAAA TCAAGACTTC CAGAGTGGGA AGAACAAATT GTATTTAATG
1301 AAAATTTTCC CTATTTGCTT CGAGGCTCTG ATGAGAGTCC TAAAGTCATC
1351 CTGTTCTTTG AGATTCTTGA TTTCTTAAGC GTGGATGAAA TTAAGAATAA
1401 TTCTGAGGTT CAAAACCAAG AATGTGGCTT TCGGAAAATT GCCTGGGCAT
1451 TTCTTAAGCT TCTGGGAGCC AATGGAAATG CAAACATCAA CTCAAAACTT
1501 CGCTTGCAGC TATATTACCC ACCTACTAAG CCTCGATCCC CATTAAGTGT
1551 TGTTGAGGCA TTTGAATGGT GGTCAAAATG TCCAAGAAAT CATTACCCAT
1601 CAACACTGTA CGTAACTGTA AGAGGACTGA AAGTTCCAGA CTGTATAAAG
1651 CCATCTTACC GCTCTATGAT GGCTCTTCAG GAGGAAAAAG GTAAACCAGT
1701 GCATTGTGAA CGTCACCATG AGTCAAGCTC AGTAGACACA GAACCTGGAT
1751 TAGAAGAGTC AAAGGAAGTA ATAAAGTGGA AACGACTCCC TGGGCAGGCT
1801 TGCCGTATCC CAAACAAACA CCTCTTCTCA CTAAATGCAG GAGAACGAGG
1851 ATGTTTTTGT CTTGATTTCT CCCACAATGG AAGAATATTA GCAGCAGCTT
1901 GTGCCAGCCG GGATGGATAT CCAATTATTT TATATGAAAT TCCTTCTGGA
1951 CGTTTCATGA GAGAATTGTG TGGCCACCTC AATATCATTT ATGATCTTTC
2001 CTGGTCAAAA GATGATCACT ACATCCTTAC TTCATCATCT GATGGCACTG
2051 CCAGGATATG GAAAAATGAA ATAAACAATA CAAATACTTT CAGAGTTTTA
2101 CCTCATCCTT CTTTTGTTTA CACGGCTAAA TTCCATCCAG CTGTAAGAGA
2151 GCTAGTAGTT ACAGGATGCT ATGATTCCAT GATACGGATA TGGAAAGTTG
2201 AGATGAGAGA AGATTCTGCC ATATTGGTCC GACAGTTTGA TGTTCACAAA
2251 AGTTTTATCA ACTCACTTTG TTTTGATACT GAAGGTCATC ATATGTATTC
2301 AGGAGATTGT ACAGGGGTGA TTGTTGTTTG GAATACCTAT GTCAAGATTA
2351 ATGATTTGGA ACATTCAGTG CACCACTGGA CTATAAATAA GGAAATTAAA
2401 GAAACTGAGT TTAAGGGAAT TCCAATAAGT TATTTGGAGA TTCATCCCAA
2451 TGGAAAACGT TTGTTAATCC ATACCAAAGA CAGTACTTTG AGAATTATGG
2501 ATCTCCGGAT ATTAGTAGCA AGGAAGTTTG TAGGAGCAGC AAATTATCGG
2551 GAGAAGATTC ATAGTACTTT GACTCCATGT GGGACTTTTC TGTTTGCTGG
2601 AAGTGAGGAT GGTATAGTGT ATGTTTGGAA CCCAGAAACA GGAGAACAAG
```

```
2651 TAGCCATGTA TTCTGACTTG CCATTCAAGT CACCCATTCG AGACATTTCT
2701 TATCATCCAT TTGAAAATAT GGTTGCATTC TGTGCATTTG GGCAAAATGA
2751 GCCAATTCTT CTGTATATTT ACGATTTCCA TGTTGCCCAG CAGGAGGCTG
2801 AAATGTTCAA ACGCTACAAT GGAACATTTC CATTACCTGG AATACACCAA
2851 AGTCAAGATG CCCTATGTAC CTGTCCAAAA CTACCCCATC AAGGCTCTTT
2901 TCAGATTGAT GAATTTGTCC ACACTGAAAG TTCTTCAACG AAGATGCAGC
2951 TAGTAAAACA GAGGCTTGAA ACTGTCACAG AGGTGATACG TTCCTGTGCT
3001 GCAAAAGTCA ACAAAAATCT CTCATTTACT TCACCACCAG CAGTTTCCTC
3051 ACAACAGTCT AAGTTAAAGC AGTCAAACAT GCTGACCGCT CAAGAGATTC
3101 TACATCAGTT TGGTTTCACT CAGACCGGGA TTATCAGCAT AGAAAGAAAG
3151 CCTTGTAACC ATCAGGTAGA TACAGCACCA ACGGTAGTGG CTCTTTATGA
3201 CTACACAGCG AATCGATCAG ATGAACTAAC CATCCATCGC GGAGACATTA
3251 TCCGAGTGTT TTTCAAAGAT AATGAAGACT GGTGGTATGG CAGCATAGGA
3301 AAGGGACAGG AAGGTTATTT TCCAGCTAAT CATGTGGCTA GTGAAACACT
3351 GTATCAAGAA CTGCCTCCTG AGATAAAGGA GCGATCCCCT CCTTTAAGCC
3401 CTGAGGAAAA AACTAAAATA GAAAAATCTC CAGCTCCTCA AAAGCAATCA
3451 ATCAATAAGA ACAAGTCCCA GGACTTCAGA CTAGGCTCAG AATCTATGAC
3501 ACATTCTGAA ATGAGAAAAG AACAGAGCCA TGAGGACCAA GGACACATAA
3551 TGGATACACG GATGAGGAAG AACAAGCAAG CAGGCAGAAA AGTCACTCTA
3601 ATAGAGTAAA GAATTGAAGA AAAGTTAAGA GCTGCCGAAA TGCACAGAGG
3651 TGAAAATGAC AAACCAAATG GAATTTCTCT TCAGAGTTCA GAATTTTCAG
3701 ATACTAAGGA GGAAGAAAGG ATCCACTACT TCTTGTTCTT ATGAATGACT
3751 CTAGAAAAAT CAGAATCAAG TTGTGGGTGG AAAAATCAAC GTGGCCTTTG
3801 AGTTCAGTTG TTATAAACCA TTGTGACTAT TGTTGGTCAA AGTATTGGTA
3851 CTTATATTGT TAGTAATTGC ATCATAATTA CATTACCAGT GTTGGAAAAC
3901 TAATGAAGAA AACACTGTAA TTGCTACTCA GCAAATGTGA ATAAAAGGTG
3951 TTTGCGTTAT TAGGATGTCT GTTAAGTAAT CATTTAATAT TATTATATTG
4001 GTAATGGTTG TATGTGTGAT GCTATGCCCA GAATATGAAG TATCTGTTTT
4051 TGAAATTCAC TTTATTTAAA AGATAAGCAG CTGACTGGGC ACGGTGCCTC
4101 ATGCCTGTAA TCCTAGCACC TTGGGAGGCT GAGGCAGGTG GATCACCTAA
4151 GGTCAGGAGT TCAACAACAC CAGCCTGACC AACATGGTGA AACCCCATCT
4201 CTACTAAAAA TACAAAAATC AGCCGGGTCT CATGGCAGGC ACCTGTAATC
4251 CCATCTACTG AGGCAGGAGA ATTGCTTGAC CCAGGAGGCA GAGGTTGCAG
4301 TGAGCCAAGA TCACGCCATT GCACTCCAGC CTGGGGGACA GAGCAAGACT
4351 CTATCTCCAA AAAACAAAAA AGATAAGCAG CTTTAGAATA TGGCGCATTC
4401 AAAACAGTCT CAGTAACAAA GACATTAAAA GAAAACAATT TACTTTCTAA
4451 TTAAAATTTT GTGTTTCTTA AGATCAAATC ATATAGGTAA CTTCATAGAC
4501 CTAAATTAAA AGTGATTTTT GGCTGGACTG GCAACAATGT TCCCAATGTC
4551 TTTACTTTTT AAAAAAGGCT TTTCATATTT AAGCACATAC CTATTTTGTA
4601 GACTTACATT GTTTAATATT TATTTTAATC TTAATATTTT TACATTATTA
4651 TATTGCATTA TTTATTTTTT CTAAGTTCCA GAATAATAGT GTCATTATTA
4701 TAGACTATAT GTTTTGAAGT TTGATATTAT AATGGGATAT TCATTTTTTG
4751 TTCTTTTCTT GACTCCTTTC TCAAGTGTGT GATAAGGTCT GCTGATAAAA
4801 TATTTAACCC CAAGAAAGTG AAAACTAATA TAAAATTAGA AAGACCTATC
4851 CAAATTAGAC AGTCAATTCC ATTAAAATAA GAAGTGAGAA AAACAATGTT
4901 GGGCATTGAG GTGTAAATTT TGCCCAGATG TATACCCAGT GTGAAATATC
4951 TTCTAATAAA AATATATTTG GCTCTTATCC CTGCACATGT AGAGGCATAA
5001 AAATTGGTAA ACATGTCCCG CTGTGTAGAA CTTTAAAAAA AAGGCATTTT
5051 TGAAAGTGTT GAGTGGCACT GATAACTGGT GAAGCCTACA GCCATCCGCC
5101 CAAAAGTCTG TTCTGATGGC ACTGAGTTTT CATTGTTCTG GATGTATAAG
5151 TCTGTGTGTC AGGTACAGCT GGGCCCAGCC AGCTTGAGTC ACTCTTGTAC
5201 AAGCTTGTTT TTTTCTGTCT TGTGAATGCA CTTGATAATT TAAAAATAAA
5251 AATATCTGTT TCTCTGCAAA AAAAAAA
```

BLAST Results

Entry HS32B1 from database EMBL:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 32B1
Score = 4445, P = 0.0e+00, identities = 889/889

Entry U93816 from database EMBL:
Human exon-trapped sequence from 6q24.
Score = 965, P = 4.0e-35, identities = 193/193

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 19 bp to 3606 bp; peptide length: 1196
Category: similarity to known protein

```
   1 MPTAESEAKV KTKVRFEKLL KTHSDLMREK KKLKKKLVRS EENISPDTIR
  51 SNLHYMKETT SDDPDTIRSN LPHIKETTSD DVSAANTNNL KKSTRVTKNK
 101 LRNTQLATEN PNGDASVEED KQGKPNKKVI KTVPQLTTQD LKPETPENKV
 151 DSTHQKTHTK PQPGVDHQKS EKANEGREET DLEEDEELMQ AYQCHVTEEM
 201 AKEIKRKIRK KLKEQLTYFP SDTLFHDDKL SSEKRKKKKE VPVFSKAETS
 251 TLTISGDTVE GEQKKESSVR SVSSDSHQDD EISSMEQSTE DSMQDDTKPK
 301 PKKTKKKTKA VADNNEDVDG DGVHEITSRD SPVYPKCLLD DDLVLGVYIH
 351 RTDRLKSDFM ISHPMVKIHV VDEHTGQYVK KDDSGRPVSS YYEKENVDYI
 401 LPIMTQPYDF KQLKSRLPEW EEQIVFNENF PYLLRGSDES PKVILFFEIL
 451 DFLSVDEIKN NSEVQNQECG FRKIAWAFLK LLGANGNANI NSKLRLQLYY
 501 PPTKPRSPLS VVEAFEWWSK CPRNHYPSTL YVTVRGLKVP DCIKPSYRSM
 551 MALQEEKGKP VHCERHHESS SVDTEPGLEE SKEVIKWKRL PGQACRIPNK
 601 HLFSLNAGER GCFCLDFSHN GRILAAACAS RDGYPIILYE IPSGRFMREL
 651 CGHLNIIYDL SWSKDDHYIL TSSSDGTARI WKNEINNTNT FRVLPHPSFV
 701 YTAKFHPAVR ELVVTGCYDS MIRIWKVEMR EDSAILVRQF DVHKSFINSL
 751 CFDTEGHHMY SGDCTGVIVV WNTYVKINDL EHSVHHWTIN KEIKETEFKG
 801 IPISYLEIHP NGKRLLIHTK DSTLRIMDLR ILVARKFVGA ANYREKIHST
 851 LTPCGTFLFA GSEDGIVYVW NPETGEQVAM YSDLPFKSPI RDISYHPFEN
 901 MVAFCAFGQN EPILLYIYDF HVAQQEAEMF KRYNGTFPLP GIHQSQDALC
 951 TCPKLPHQGS FQIDEFVHTE SSSTKMQLVK QRLETVTEVI RSCAAKVNKN
1001 LSFTSPPAVS SQQSKLKQSN MLTAQEILHQ FGFTQTGIIS IERKPCNHQV
1051 DTAPTVVALY DYTANRSDEL TIHRGDIIRV FFKDNEDWWY GSIGKGQEGY
1101 FPANHVASET LYQELPPEIK ERSPPLSPEE KTKIEKSPAP QKQSINKNKS
1151 QDFRLGSESM THSEMRKEQS HEDQGHIMDT RMRKNKQAGR KVTLIE
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_1n3, frame 1

TREMBL:U92792_1 gene: "tup1"; product: "Tup1"; Schizosaccharomyces pombe general transcriptional repressor Tup1 (tup1) mRNA, complete cds., N = 1, Score = 186, P = 1e-10

TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

PIR:T02533 hypothetical protein F13M22.17 - Arabidopsis thaliana, N = 2, Score = 228, P = 1e-13

TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

TREMBL:CER03E1_1 gene: "R03E1.1"; Caenorhabditis elegans cosmid R03E1, N = 1, Score = 215, P = 2.3e-13

SWISSPROT:YZLL_CAEEL HYPOTHETICAL 43.1 KD TRP-ASP REPEATS CONTAINING PROTEIN K04G11.4 IN CHROMOSOME X., N = 1, Score = 203, P = 7.1e-13

```
>TREMBL:AF104258_1 gene: "Pmc733"; product: "putative copper-inducible 35.6
     kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein
     (Pmc733) mRNA, complete cds.
            Length = 321

  HSPs:

 Score = 235 (35.3 bits), Expect = 4.6e-18, P = 4.6e-18
 Identities = 59/225 (26%), Positives = 111/225 (49%)

Query:   647 MRELCGHLNIIYDLSWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKFH 706
             + E  GH + I DLSWSK+   +L++S D T R+W  ++   +  +V  H ++V   +F+
Sbjct:    63 VHEFYGHGDAILDLSWSKNGD-LLSASMDKTVRLW--QVGRDSCLKVFSHTNYVTCVQFN 119

Query:   707 PAVRELVVTGCYDSMIRIWKVEMREDSAILVRQFDVHKSFINSLCFDTEGHHMYSGDCTG 766
             P      +TGC D ++RIW V          LV  +   K  + ++C+  +G   +G  TG
Sbjct:   120 PTNGNYFITGCIDGLVRIWDVRK-----CLVVDWANSKEIVTAVCYRPDGKGAVAGTITG 174

Query:   767 VIVVWNTYVKINDLEHSVHHWTINKEIKETEFKGIPISYLEIHPNGKRLLIHTKDSTLRI 826
                 ++      +LE  V    ++N   K      + +   Y     P  K+L++ + D+ +RI
```

```
Sbjct:   175 NCRYYDASENRLELESQV---SLNGRKKSLHKRIVGFQYCPSDP--KKLMVTSGDAQVRI 229

Query:   827 MDLRILVARKFVGAANYREKIHSTLTPCGTFLFAGSEDGIVYVWN 871
             +D   +++  + G  +    ++  + TP G  + + S+D  +Y+WN
Sbjct:   230 LDGAHVISN-YKGLQS-SSQVARSFTPDGDHIVSASDDSRIYMWN 272
```

Pedant information for DKFZphtes3_1n3, frame 1

Report for DKFZphtes3_1n3.1

```
[LENGTH]        1196
[MW]            137114.70
[pI]            6.79
[HOMOL]         SWISSPROT:YKY4_CAEEL HYPOTHETICAL 40.4 KD TRP-ASP REPEATS CONTAINING PROTEIN
C14B1.4 IN CHROMOSOME III. 8e-21
[FUNCAT]        99 unclassified proteins     [S. cerevisiae, YKL121w] 2e-11
[FUNCAT]        04.05.01.01 general transcription activities        [S. cerevisiae, YBR198c
TAF90 - TFIID subunit] 4e-10
[FUNCAT]        30.10 nuclear organization    [S. cerevisiae, YBR198c TAF90 - TFIID subunit]
4e-10
[FUNCAT]        06.10 assembly of protein complexes  [S. cerevisiae, YPR178w] 1e-08
[FUNCAT]        04.05.03 mrna processing (splicing)  [S. cerevisiae, YPR178w] 1e-08
[FUNCAT]        03.22 cell cycle control and mitosis [S. cerevisiae, YDR364c] 4e-08
[FUNCAT]        03.16 dna synthesis and replication  [S. cerevisiae, YDR364c] 4e-08
[FUNCAT]        08.07 vesicular transport (golgi network, etc.)     [S. cerevisiae, YDL145c]
9e-08
[FUNCAT]        30.09 organization of intracellular transport vesicles     [S. cerevisiae,
YDL145c] 9e-08
[FUNCAT]        04.05.01.04 transcriptional control  [S. cerevisiae, YCR084c] 2e-07
[FUNCAT]        10.99 other signal-transduction activities  [S. cerevisiae, YHL002w] 7e-07
[FUNCAT]        98 classification not yet clear-cut  [S. cerevisiae, YFR024c-a] 2e-06
[FUNCAT]        02.16 fermentation     [S. cerevisiae, YMR116c] 4e-06
[FUNCAT]        30.03 organization of cytoplasm      [S. cerevisiae, YMR116c] 4e-06
[FUNCAT]        05.04 translation (initiation, elongation and termination)  [S. cerevisiae,
YMR116c] 4e-06
[FUNCAT]        03.10 sporulation and germination     [S. cerevisiae, YFL009w] 4e-05
[FUNCAT]        03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL009w]
4e-05
[FUNCAT]        30.04 organization of cytoskeleton    [S. cerevisiae, YFL009w] 4e-05
[FUNCAT]        03.01 cell growth      [S. cerevisiae, YCR088w] 6e-05
[FUNCAT]        03.25 cytokinesis      [S. cerevisiae, YCR057c] 7e-05
[BLOCKS]        BL00024H
[SCOP]          d1tbgd_ 2.46.3.1.1 beta1-subunit of the signal-transducing  3e-91
[SCOP]          d1gfc__ 2.21.2.1.9 Growth factor receptor-bound protein 2 (GRB2), N 4e-14
[SCOP]          d1fmk_1 2.21.2.1.8 (1-64) c-src tyrosine kinase [human (Hom 5e-15
[SCOP]          d1ad5b1 2.21.2.1.7 (1-63) Hemapoetic cell kinase Hck [human (Hom 3e-15
[SCOP]          d1lcka1 2.21.2.1.16 (1-54) p56-lck tyrosine kinase, SH3 domain [huma 1e-13
[SCOP]          d1qwea_ 2.21.2.1.15 Src kinase, SH3 domain [Avian sarcoma virus 2e-15
[SCOP]          d1shg__ 2.21.2.1.6 alpha-Spectrin, SH3 domain [chicken (Gallu 2e-13
[SCOP]          d1prmc_ 2.21.2.1.13 Src kinase, SH3 domain [chicken (Gallus gallus) 2e-15
[SCOP]          d1hsq__ 2.21.2.1.12 Phospholipase C, SH3 domain [human (Hom 2e-13
[SCOP]          d1aboa_ 2.21.2.1.3 Abl tyrosine kinase, SH3 domain [Mouse (Mu 3e-13
[SCOP]          d1efna_ 2.21.2.1.2 Fyn, SH3 domain [human (Homo sapiens) 2e-15
[SCOP]          d1sema_ 2.21.2.1.11 Growth factor receptor-bound protein 2 (GRB2), N 1e-13
[SCOP]          d1gbqa_ 2.21.2.1.10 Growth factor receptor-bound protein 2 (GRB2), N 3e-16
[SCOP]          d1ckaa_ 2.21.2.1.1 C-Crk, N-terminal SH3 domain [mouse (Mu 3e-15
[EC]            3.1.4.3 Phospholipase C 2e-07
[EC]            3.1.4.11 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase 7e-07
[EC]            3.6.1.32 Myosin ATPase 7e-07
[EC]            2.7.1.112 Protein-tyrosine kinase 8e-06
[PIRKW]         nucleus 2e-08
[PIRKW]         phosphotransferase 8e-06
[PIRKW]         plasma 4e-07
[PIRKW]         duplication 4e-07
[PIRKW]         phosphoric diester hydrolase 2e-07
[PIRKW]         tandem repeat 7e-07
[PIRKW]         hormone 4e-07
[PIRKW]         transmembrane protein 2e-06
[PIRKW]         stomach 4e-07
[PIRKW]         actin binding 7e-07
[PIRKW]         ATP 7e-07
[PIRKW]         phosphoprotein 7e-07
[PIRKW]         signal transduction 7e-09
[PIRKW]         heterotrimer 7e-09
[PIRKW]         P-loop 7e-07
[PIRKW]         hydrolase 7e-07
[PIRKW]         transcription regulation 5e-06
[PIRKW]         GTP binding 7e-09
```

```
[SUPFAM]        1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase II 2e-07
[SUPFAM]        SH3 homology 2e-07
[SUPFAM]        SH2 homology 2e-07
[SUPFAM]        protozoan myosin heavy chain IB 7e-07
[SUPFAM]        myosin motor domain homology 7e-07
[SUPFAM]        pleckstrin repeat homology 2e-07
[SUPFAM]        protein-tyrosine kinase src 8e-06
[SUPFAM]        WD repeat homology 3e-12
[SUPFAM]        1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain Y homology 2e-
07
[SUPFAM]        protein kinase homology 8e-06
[SUPFAM]        1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X homology 2e-
07
[SUPFAM]        GTP-binding regulatory protein beta chain 7e-09
[SUPFAM]        yeast coatomer complex alpha chain 4e-07
[PROSITE]       RGD     1
[PROSITE]       MYRISTYL        6
[PROSITE]       AMIDATION       2
[PROSITE]       CAMP_PHOSPHO_SITE       4
[PROSITE]       CK2_PHOSPHO_SITE        25
[PROSITE]       TYR_PHOSPHO_SITE        4
[PROSITE]       PKC_PHOSPHO_SITE        19
[PROSITE]       ASN_GLYCOSYLATION       6
[PFAM]          Src homology domain 3
[PFAM]          WD domain, G-beta repeats
[KW]            Irregular
[KW]            3D
[KW]            LOW_COMPLEXITY          5.77 %
[KW]            COILED_COIL             2.42 %

SEQ    MPTAESEAKVKTKVRFEKLLKTHSDLMREKKKLKKKLVRSEENISPDTIRSNLHYMKETT
SEG    .............................xxxxxxxx.......................
COILS  ................CCCCCCCCCCCCCCCCCCCCCCCCCCCCC...............
1gotB  ............................................................

SEQ    SDDPDTIRSNLPHIKETTSDDVSAANTNNLKKSTRVTKNKLRNTQLATENPNGDASVEED
SEG    ............................................................
COILS  ............................................................
1gotB  ............................................................

SEQ    KQGKPNKKVIKTVPQLTTQDLKPETPENKVDSTHQKTHTKPQPGVDHQKSEKANEGREET
SEG    .........................................................xxx
COILS  ............................................................
1gotB  ............................................................

SEQ    DLEEDEELMQAYQCHVTEEMAKEIKRKIRKKLKEQLTYFPSDTLFHDDKLSSEKRKKKKE
SEG    xxxxxxxx.............xxxxxxxxxxxxxxx............xxxxxxxxxxxx
COILS  ............................................................
1gotB  ............................................................

SEQ    VPVFSKAETSTLTISGDTVEGEQKKESSVRSVSSDSHQDDEISSMEQSTEDSMQDDTKPK
SEG    ..........................xxxxxxxxxx....................xxxx
COILS  ............................................................
1gotB  ............................................................

SEQ    PKKTKKKTKAVADNNEDVDGDGVHEITSRDSPVYPKCLLDDDLVLGVYIHRTDRLKSDFM
SEG    xxxxxxxxx...................................................
COILS  ............................................................
1gotB  ............................................................

SEQ    ISHPMVKIHVVDEHTGQYVKKDDSGRPVSSYYEKENVDYILPIMTQPYDFKQLKSRLPEW
SEG    ............................................................
COILS  ............................................................
1gotB  ............................................................

SEQ    EEQIVFNENFPYLLRGSDESPKVILFFEILDFLSVDEIKNNSEVQNQECGFRKIAWAFLK
SEG    ............................................................
COILS  ............................................................
1gotB  ............................................................

SEQ    LLGANGNANINSKLRLQLYYPPTKPRSPLSVVEAFEWWSKCPRNHYPSTLYVTVRGLKVP
SEG    ............................................................
COILS  ............................................................
1gotB  ............................................................

SEQ    DCIKPSYRSMMALQEEKGKPVHCERHHESSSVDTEPGLEESKEVIKWKRLPGQACRIPNK
SEG    ............................................................
COILS  ............................................................
1gotB  ............................................................
```

SEQ	HLFSLNAGERGCFCLDFSHNGRILAAACASRDGYPIILYEIPSGRFMRELCGHLNIIYDL
SEG	....
COILS	....
1gotB	....CEEEEEECCCCCEEEE
SEQ	SWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKFHPAVRELVVTGCYDS
SEG	....
COILS	....
1gotB	EETTTTTEEEEEETTTEEEEEETT--TTCEEEEEETTTCEEEEEETTT-TCEEEEEETTT
SEQ	MIRIWKVEMREDSAILVRQFDVHKSFINSLCFDTEGHHMYSGDCTGVIVVWNTYVKINDL
SEG	....
COILS	....
1gotB	EEEEEETTTTTBTTEEEEEEEECCCCCE-EEEEEEETTEEEEEEETTTEEEEEE......
SEQ	EHSVHHWTINKEIKETEFKGIPISYLEIHPNGKRLLIHTKDSTLRIMDLRILVARKFVGA
SEG	....
COILS	....
1gotB	....
SEQ	ANYREKIHSTLTPCGTFLFAGSEDGIVYVWNPETGEQVAMYSDLPFKSPIRDISYHPFEN
SEG	....
COILS	....
1gotB	....
SEQ	MVAFCAFGQNEPILLYIYDFHVAQQEAEMFKRYNGTFPLPGIHQSQDALCTCPKLPHQGS
SEG	....
COILS	....
1gotB	....
SEQ	FQIDEFVHTESSSTKMQLVKQRLETVTEVIRSCAAKVNKNLSFTSPPAVSSQQSKLKQSN
SEG	....
COILS	....
1gotB	....
SEQ	MLTAQEILHQFGFTQTGIISIERKPCNHQVDTAPTVVALYDYTANRSDELTIHRGDIIRV
SEG	....
COILS	....
1gotB	....
SEQ	FFKDNEDWWYGSIGKGQEGYFPANHVASETLYQELPPEIKERSPPLSPEEKTKIEKSPAP
SEG	....
COILS	....
1gotB	....
SEQ	QKQSINKNKSQDFRLGSESMTHSEMRKEQSHEDQGHIMDTRMRKNKQAGRKVTLIE
SEG	....
COILS	....
1gotB	....

Prosite for DKFZphtes3_1n3.1

PS00001	460->464	ASN_GLYCOSYLATION	PDOC00001
PS00001	686->690	ASN_GLYCOSYLATION	PDOC00001
PS00001	934->938	ASN_GLYCOSYLATION	PDOC00001
PS00001	1000->1004	ASN_GLYCOSYLATION	PDOC00001
PS00001	1065->1069	ASN_GLYCOSYLATION	PDOC00001
PS00001	1148->1152	ASN_GLYCOSYLATION	PDOC00001
PS00004	91->95	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	264->268	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	305->309	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1190->1194	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00005	93->96	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	304->307	PKC_PHOSPHO_SITE	PDOC00005
PS00005	327->330	PKC_PHOSPHO_SITE	PDOC00005
PS00005	352->355	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	440->443	PKC_PHOSPHO_SITE	PDOC00005
PS00005	533->536	PKC_PHOSPHO_SITE	PDOC00005
PS00005	546->549	PKC_PHOSPHO_SITE	PDOC00005
PS00005	643->646	PKC_PHOSPHO_SITE	PDOC00005
PS00005	677->680	PKC_PHOSPHO_SITE	PDOC00005
PS00005	690->693	PKC_PHOSPHO_SITE	PDOC00005
PS00005	702->705	PKC_PHOSPHO_SITE	PDOC00005

PS00005	823->826	PKC_PHOSPHO_SITE	PDOC00005
PS00005	973->976	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00006	283->287	CK2_PHOSPHO_SITE	PDOC00006
PS00006	288->292	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	327->331	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	454->458	CK2_PHOSPHO_SITE	PDOC00006
PS00006	510->514	CK2_PHOSPHO_SITE	PDOC00006
PS00006	570->574	CK2_PHOSPHO_SITE	PDOC00006
PS00006	663->667	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	985->989	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1023->1027	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1127->1131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1132->1136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1161->1165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1170->1174	CK2_PHOSPHO_SITE	PDOC00006
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	210->219	TYR_PHOSPHO_SITE	PDOC00007
PS00008	483->489	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	716->722	MYRISTYL	PDOC00008
PS00008	800->806	MYRISTYL	PDOC00008
PS00008	861->867	MYRISTYL	PDOC00008
PS00008	941->947	MYRISTYL	PDOC00008
PS00009	811->815	AMIDATION	PDOC00009
PS00009	1188->1192	AMIDATION	PDOC00009
PS00016	1074->1077	RGD	PDOC00016

Pfam for DKFZphtes3_1n3.1

```
HMM_NAME     WD domain, G-beta repeats

HMM              *MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*
                  + GH+N ++++++S D ++ I+++S DGT R+W
Query      650   LCGHLNIIYDLSWSKDDHY-ILTSSSDGTARIWK    682


HMM_NAME     Src homology domain 3

HMM              *pyVIALYDYqAqdpDELSFkEGDIIiIIEdsDD.WWrgRnnnTNGQEGW
                  P+V+ALYDY+A+++DEL++ +GDII +  ++++ WW+G       GQEG+
Query     1054   PTVVALYDYTANRSDELTIHRGDIIRVFFKDNEDWWYGSIGK--GQEGY  1100

HMM              IPSNYVEPi*
                 +P+N V+ +
Query     1101   FPANHVASE   1109
```

DKFZphtes3_20c21
----

group: testes derived

DKFZphtes3_20c21 encodes a novel 708 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /map="22q11.2-12.2"

Insert length: 3997 bp
Poly A stretch at pos. 3877, polyadenylation signal at pos. 3853

```
   1 GGTAGGCGGG GCGGCGCGTG ACCTAAGGCC TCTCTGCCGC GCGCGCAGGT
  51 ACGGGGCAGA AGTCGCAGGT ACCCAGCTGC TGCCCACGTT TCTGGTCCAG
 101 AGTCCCGAAC CCCGAGCACT GGGATGCCTG GCTACTCCGA GCCAAGGCAC
 151 TGATGTTTGA ACTGGAAACT TCAAAACGTT TAATAAGAGT CTTCAGGATG
 201 GGTTTGAACT AGACAAGCTA GAAATTTCTT TAGAACACCA GCTCTAGCAT
 251 GCATCTCCCA CTTTTGGCTT TCCTGGAGAG GAGCTTGAAG AGGTGGTTCT
 301 GCAGACAGCC ACAGTGATAC TCAGGAAACC AGAGGAATGG ATTTGACTTT
 351 TCTGCTAGGA TTCTTTGTTA TAGTTTCTCC CTGAGTTGTA AGAGGCATGG
 401 AAATATACAT GAAACTGAAG AACCTGCAAG GAAGGGAAGT GGAACTTTCC
 451 ATGCTGAGTG AAAACTAACC AAGTGGCAGT TGTGACTGAA AACACTGAAA
 501 CCTACCACGT CCAGATTCAC TGGATTGGGG GATAGAGGAA CGGTCACAGC
 551 TAGGGAGAAA GAAGTGATAC CGGAAAAGAA AACCTAAATG AAGAGAATGA
 601 GGATGACTGC ACAGTAGATG GCCACCTCTA CCTCCACAGA GGCAAAGTCA
 651 GCCTCGTGGT GGAATTATTT TTTTCTTTAT GATGGTTCCA AGGTAAAGGA
 701 AGAAGGCGAT CCAACAAGAG CTGGCATTTG TTACTTTTAT CCTTCCCAGA
 751 CCCTGCTAGA CCAACAGGAG TTGCTTTGTG GACAGATTGC TGGAGTTGTC
 801 CGCTGTGTTT CTGACATTTC TGACTCTCCT CCTACTCTTG TTCGTCTGAG
 851 AAAACTGAAG TTTGCCATAA AAGTTGATGG AGATTACCTT TGGGTGCTGG
 901 GCTGTGCTGT GGAGCTCCCT GATGTCAGCT GCAAGCGGTT TCTGGATCAG
 951 CTAGTTGGAT TCTTTAATTT TTACAATGGA CCTGTTTCCC TAGCTTATGA
1001 GAACTGTTCT CAGGAAGAAC TGAGCACGGA GTGGGACACC TTCATCGAGC
1051 AAATTCTGAA AAACACCAGT GATCTGCATA AGATTTTCAA TTCCCTCTGG
1101 AACTTGGACC AAACTAAAGT GGAGCCCCTG TTGTTGCTGA AGGCAGCCCG
1151 CATTCTGCAG ACCTGCCAGC GCTCGCCTCA CATTCTCGCT GGCTGCATCC
1201 TCTATAAAGG ACTGATTGTC AGCACCCAAC TCCCGCCCTC CCTCACCGCC
1251 AAGGTCCTGC TTCACCGAAC AGCACCTCAG GAGCAGAGAC TCCCTACGGG
1301 AGGGGATGCC CCGCAGGAAC ATGGAGCGGC ATTGCCCCCG AATGTCCAGA
1351 TTATCCCTGT TTTTGTGACC AAAGAGGAAG CCATTAGTCT CCACGAGTTC
1401 CCGGTGGAAC AGATGACAAG GTCTCTAGCA TCTCCAGCAG GACTCCAGGA
1451 TGGTTCAGCC CAGCACCATC CAAAGGGTGG GAGCACATCT GCCCTGAAAG
1501 AAAACGCCAC TGGCCATGTG GAATCCATGG CCTGGACCAC CCCAGATCCC
1551 ACATCCCCTG ACGAAGCTTG TCCAGATGGC AGGAAGGAGA ACGGATGCTT
1601 GTCTGGCCAT GATCTGGAGA GCATCAGGCC CGCAGGACTG CACAACTCTG
1651 CCAGGGGTGA GGTTCTTGGC CTCAGCTCCT CCCTGGGGAA GGAACTAGTC
1701 TTTCTCCAAG AAGAACTCGA CTTGTCTGAA ATCCACATTC CAGAGGCTCA
1751 GGAAGTGGAA ATGGCCTCAG GTCATTTTGC CTTCCTACAT GTGCCTGTTC
1801 CAGATGGCAG GGCTCCTTAC TGCAAGGCAT CTCTCAGCGC CTCCAGCAGC
1851 CTGGAACCCA CGCCTCCTGA GGACACAGCC ATCAGCAGCT TGCGCCCTCC
1901 CTCTGCTCCT GAGATGCTGA CCCAGCATGG AGCCCAAGAG CAGGTCGAAG
1951 ACCATCCTGG CCATAGCAGC CAAGCCCCCA TTCCCAGAGC AGACCCTCTC
2001 CCCAGAAGGA CCCGCAGGCC CTTGTTATTG CCTCGCTTAG ATCCAGGACA
2051 GAGAGGAAAC AAGCTTCCCA CGGGGGAACA AGGCCTGGAT GAGGATGTTG
2101 ATGGGGTCTG TGAAAGCCAC GCAGCCCCTG GTCTGGAATG CAGTTCAGGC
2151 TCAGCAAACT GTCAGGGTGC TGGCCCCTCT GCAGATGGAA TCAGCTCCAG
2201 GCTGACACCA GCAGAGTCCT GCATGGGGCT CGTGAGGATG AATCTCTACA
2251 CTCACTGCGT CAAAGGGCTG ATGCTGTCCC TGCTGGCTGA GGAGCCGCTG
2301 CTGGGAGACA GCGCAGCCAT AGAGGAAGTG TACCACAGCA GCCTGGCTTC
2351 ACTGAATGGG CTGGAAGTCC ACCTGAAAGA GACGCTGCCC AGGGATGAGG
2401 CAGCCTCCAC GAGCAGCACC TACAACTTCA CATATTACGA CCGCATTCAG
2451 AGCTTGCTGA TGGCAAACCT GCCGCAGGTG GCCACCCCGC ATGATCGCCG
2501 CTTCCTCCAG GCCGTCAGCC TGATGCATAG CGAATTTGCC CAGCTGCCCG
2551 CGCTTTATGA AATGACTGTC AGAAATGCCT CCACGGCTGT GTACGCCTGT
2601 TGCAACCCCA TCCAGGAGAC ATATTTCCAG CAGCTGGCAC CTGCAGCACG
2651 GAGCTCCGGC TTCCCAAACC CTCAGGATGG CGCCTTCAGC CTCTCCGGCA
2701 AAGCAAAGCA GAAGCTGCTG AAGCACGGGG TGAACTTGCT CTGAACTGCA
2751 CCCAGGAGGT GACTGGGAAG GAGAAAACCA GCAAAGGAAG CTCTGCCTTT
2801 TATAATTGAA AAGGCCCCTC TATTTTATTT TTCTTGAAAA CATTCCCTTT
```

```
2851 TTTAGGAACC AAATGATATT TGAGTTTTTG TTATTCCTTT TGCAGATTGG
2901 GATGTGTTTT GGGGGCAGGG GTTAGTTCTT CAGGTCGGCA GACCCAGAGC
2951 ACTTGATAAA GAACTGTATT TAATCGGTAG TGTTGGGGCC GGGACGGGCT
3001 TGGCTCCCTC TCTGCCATAC TGAGCCTGAG GTATTTCATA TCTCCTGCTG
3051 TTCCATCCCA GCTTGAATTG GTGCCACAAG CTTCCAAGTT GGCATTTTTT
3101 CTAGAACCTG ATCGTCCACT AGCCCAGAGT GTGTGTGTTC AACCCCCACA
3151 CCAGGTGGTG GTAGGCGGTG TGACTGCACA GCGAGGTGCC GGATCTGTGA
3201 GCAGGCCGAC TCCACTCCCA CGCCGCAGGT AGGTTTCTCC AGTGCGCTCT
3251 TGCTGGGAGG TCCGGATCGT TCCTGCAGGG AAGCGGCAGC ACACGGAGAC
3301 CACTTGGTTG AATTCTGTTG GAACTCTACT CAAATCTAGG GGCGTCTTCT
3351 TTGGACCCAC AATGGGGGCA AGCCTTAATA ATATGGAAGG GAGTTTGGGC
3401 TTTAGAGATC CCTTTATAAA AGCTCTGGGG GCTGAGCCCT GAGAATTCAG
3451 TGACAACAGG ACCAACCTGC GCTGCCTTTG ACTACAAGTG GGCCGTGCAG
3501 CTGGTTCCTC TCGAGCGAGT GTCCCTAAAT AGGAGTTTAC AAGATGTCTG
3551 GGGGTAAAAG CACTGTGCTT TTCAGTGGTG GCTGCGTGAA AGGGAGCGAC
3601 ACTCAGCTGT GTGTTCCTGG GCTTGTGTGG TACTTAGAAC CTCAGTTCTA
3651 TTACGTTATA GTCAGACATT TTTTTGACAG TATGAGACAG ACTGCAGGAT
3701 GAAAATATTT GTCAAAATCT TAACTGAATG TTTACTGGAA GTACTTGAGA
3751 TTCCATTTGA GAGTTGTATT GTTAATAATT TCATGTCAGT GAACTGATAT
3801 CTGATGTTTA TGATATGGTG TCTTTTTCTT GAAACAAGCT TCCAAGGGCT
3851 AGAAATAAAA TAGCCAAAAA ATGCTGGAAA AAAAAAAAAA AAAAAAAAAA
3901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3951 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA
```

BLAST Results

Entry HS1048E9 from database EMBLNEW:
Human DNA sequence from clone 1048E9 on chromosome 22q11.2-12.2
Contains pseudogene similar to ribosomal protein S3A and part of a gene
similar to C.elegans protein CE02118, ESTs, STS, GSS.
Score = 6540, P = 0.0e+00, identities = 1308/1308
~14 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 618 bp to 2741 bp; peptide length: 708
Category: putative protein
Classification: no clue

```
  1 MATSTSTEAK SASWWNYFFL YDGSKVKEEG DPTRAGICYF YPSQTLLDQQ
 51 ELLCGQIAGV VRCVSDISDS PPTLVRLRKL KFAIKVDGDY LWVLGCAVEL
101 PDVSCKRFLD QLVGFFNFYN GPVSLAYENC SQEELSTEWD TFIEQILKNT
151 SDLHKIFNSL WNLDQTKVEP LLLLKAARIL QTCQRSPHIL AGCILYKGLI
201 VSTQLPPSLT AKVLLHRTAP QEQRLPTGGD APQEHGAALP PNVQIIPVFV
251 TKEEAISLHE FPVEQMTRSL ASPAGLQDGS AQHHPKGGST SALKENATGH
301 VESMAWTTPD PTSPDEACPD GRKENGCLSG HDLESIRPAG LHNSARGEVL
351 GLSSSLGKEL VFLQEELDLS EIHIPEAQEV EMASGHFAFL HVPVPDGRAP
401 YCKASLSASS SLEPTPPEDT AISSLRPPSA PEMLTQHGAQ EQVEDHPGHS
451 SQAPIPRADP LPRRTRRPLL LPRLDPGQRG NKLPTGEQGL DEDVDGVCES
501 HAAPGLECSS GSANCQGAGP SADGISSRLT PAESCMGLVR MNLYTHCVKG
551 LMLSLLAEEP LLGDSAAIEE VYHSSLASLN GLEVHLKETL PRDEAASTSS
601 TYNFTYYDRI QSLLMANLPQ VATPHDRRFL QAVSLMHSEF AQLPALYEMT
651 VRNASTAVYA CCNPIQETYF QQLAPAARSS GFPNPQDGAF SLSGKAKQKL
701 LKHGVNLL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_20c21, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_20c21, frame 3

Report for DKFZphtes3_20c21.3

```
[LENGTH]        708
[MW]            76900.23
[pI]            5.30
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY      6.36 %

SEQ    MATSTSTEAKSASWWNYFFLYDGSKVKEEGDPTRAGICYFYPSQTLLDQQELLCGQIAGV
SEG    .xxxxxxxxxxxx...............................................
PRD    cccccccccccccccceeeeeccccccccccccccccceeeecchhhhhhhhhhhcccee

SEQ    VRCVSDISDSPPTLVRLRKLKFAIKVDGDYLWVLGCAVELPDVSCKRFLDQLVGFFNFYN
SEG    ............................................................
PRD    eeeeeecccccchhhhhhhhheeeecccceeeeeeeeeeccccchhhhhhhhheeeecc

SEQ    GPVSLAYENCSQEELSTEWDTFIEQILKNTSDLHKIFNSLWNLDQTKVEPLLLLKAARIL
SEG    ............................................................
PRD    cccccccccchhhhhhhhhhhhhhhhhhcchhhhhhhcccccccccchhhhhhhhhhh

SEQ    QTCQRSPHILAGCILYKGLIVSTQLPPSLTAKVLLHRTAPQEQRLPTGGDAPQEHGAALP
SEG    ............................................................
PRD    hhhhcccchhhhhhhcccccccccccchhhhhhhhhcccccccccccccccccccccc

SEQ    PNVQIIPVFVTKEEAISLHEFPVEQMTRSLASPAGLQDGSAQHHPKGGSTSALKENATGH
SEG    ............................................................
PRD    ccceeeeeeeeccccceeeccccchhhhhhhcccccccccccccccccchhhhhhhcccc

SEQ    VESMAWTTPDPTSPDEACPDGRKENGCLSGHDLESIRPAGLHNSARGEVLGLSSSLGKEL
SEG    ............................................................
PRD    cccccccccccccccccccccccccccccccccccccccccccccceeeeeeccccchhh

SEQ    VFLQEELDLSEIHIPEAQEVEMASGHFAFLHVPVPDGRAPYCKASLSASSSLEPTPPEDT
SEG    ............................................................
PRD    hhhhhhhcccccccccchhhhhhccceeeeeeeccccccceeeccccccccccccccccc

SEQ    AISSLRPPSAPEMLTQHGAQEQVEDHPGHSSQAPIPRADPLPRRTRRPLLLPRLDPGQRG
SEG    ...................................xxxxxxxxxxxxxxxxxxxxx....
PRD    ccccccccchhhhhhcccccceeecccccccccccccccccccccccccccccccccccc

SEQ    NKLPTGEQGLDEDVDGVCESHAAPGLECSSGSANCQGAGPSADGISSRLTPAESCMGLVR
SEG    ............................................................
PRD    ccccccccccccccccccccccccccccccccccccccccccccccccccccceeeee

SEQ    MNLYTHCVKGLMLSLLAEEPLLGDSAAIEEVYHSSLASLNGLEVHLKETLPRDEAASTSS
SEG    ..........xxxxxxxxxxxx......................................
PRD    ceeeeeeehhhhhhhhhcccccchhhhhhhhhhcccccchhhhhhhcccccccccccc

SEQ    TYNFTYYDRIQSLLMANLPQVATPHDRRFLQAVSLMHSEFAQLPALYEMTVRNASTAVYA
SEG    ............................................................
PRD    ccceeeehhhhhhhhhcccccccccchhhhhhhhhhhhhhhcchhhhhhhhhhccceeeee

SEQ    CCNPIQETYFQQLAPAARSSGFPNPQDGAFSLSGKAKQKLLKHGVNLL
SEG    ................................................
PRD    ecccchhhhhhhhhhhhhhhcccccccccceeeecchhhhhhhhhhccccc
```

(No Prosite data available for DKFZphtes3_20c21.3)

(No Pfam data available for DKFZphtes3_20c21.3)

DKFZphtes3_20k2
----

group: signal transduction

DKFZphtes3_20k2 encodes a novel 839 amino acid protein with strong similarity to rat vanilloid receptor subtype 1.

VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. The novel protein is the human orthologue of rat VR1.

The new protein can find application as a target for the development of new nociception-modulating drugs.

strong similarity to rat vanilloid receptor subtype 1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4187 bp
Poly A stretch at pos. 4154, polyadenylation signal at pos. 4135

```
   1 GGCTCAGGCA GGCCTGGCCC AGAGTCACGC TGGCAACCAC GAGTTTGGGA
  51 AGCAGTCGTA TTCTCTCTCT CTCTCTCTCT CTCTCAGTAT CCATGACAGT
 101 GTGATGGAGA GTCTCTGCCG TGCCATCTGG GATGCAAACC GTCCCTGTGT
 151 CCCCCACGTC CAGGCCGTAG ATGCTCCCCG CCGGTCAGTC ACTTAGTCGT
 201 CAGATCGCCC GTCCTGGTAT CACAGTGCTT CTGTTCAGGT TGCACACTGG
 251 GCCACAGAGG ATCCAGCAAG GATGAAGAAA TGGAGCAGCA CAGACTTGGG
 301 GGCAGCTGCG GACCCACTCC AAAAGGACAC CTGCCCAGAC CCCCTGGATG
 351 GAGACCCTAA CTCCAGGCCA CCTCCAGCCA AGCCCCAGCT CTCCACGGCC
 401 AAGAGCCGCA CCCGGCTCTT TGGGAAGGGT GACTCGGAGG AGGCTTTCCC
 451 GGTGGATTGC CCTCACGAGG AAGGTGAGCT GGACTCCTGC CCGACCATCA
 501 CAGTCAGCCC TGTTATCACC ATCCAGAGGC CAGGAGACGG CCCCACCGGT
 551 GCCAGGCTGC TGTCCCAGGA CTCTGTCGCC GCCAGCACCG AGAAGACCCT
 601 CAGGCTCTAT GATCGCAGGA GTATCTTTGA AGCCGTTGCT CAGAATAACT
 651 GCCAGGATCT GGAGAGCCTG CTGCTCTTCC TGCAGAAGAG CAAGAAGCAC
 701 CTCACAGACA ACGAGTTCAA AGACCCTGAG ACAGGGAAGA CCTGTCTGCT
 751 GAAAGCCATG CTCAACCTGC ATGACGGACA GAACACCACC ATCCCCCTGC
 801 TCCTGGAGAT CGCGCGGCAA ACGGACAGCC TGAAGGAGCT TGTCAACGCC
 851 AGCTACACGG ACAGCTACTA CAAGGGCCAG ACAGCACTGC ACATCGCCAT
 901 CGAGAGACGC AACATGGCCC TGGTGACCCT CCTGGTGGAG AACGGAGCAG
 951 ACGTCCAGGC TGCGGCCCAT GGGGACTTCT TTAAGAAAAC CAAAGGGCGG
1001 CCTGGATTCT ACTTCGGTGA ACTGCCCCTG TCCCTGGCCG CGTGCACCAA
1051 CCAGCTGGGC ATCGTGAAGT TCCTGCTGCA GAACTCCTGG CAGACGGCCG
1101 ACATCAGCGC CAGGGACTCG GTGGGCAACA CGGTGCTGCA CGCCCTGGTG
1151 GAGGTGGCCG ACAACACGGC CGACAACACG AAGTTTGTGA CGAGCATGTA
1201 CAATGAGATT CTGATCCTGG GGGCCAAACT GCACCCGACG CTGAAGCTGG
1251 AGGAGCTCAC CAACAAGAAG GGAATGACGC CGCTGGCTCT GGCAGCTGGG
1301 ACCGGGAAGA TCGGGGTCTT GGCCTATATT CTCCAGCGGG AGATCCAGGA
1351 GCCCGAGTGC AGGCACCTGT CCAGGAAGTT CACCGAGTGG GCCTACGGGC
1401 CCGTGCACTC CTCGCTGTAC GACCTGTCCT GCATCGACAC CTGCGAGAAG
1451 AACTCGGTGC TGGAGGTGAT CGCCTACAGC AGCAGCGAGA CCCCTAATCG
1501 CCACGACATG CTCTTGGTGG AGCCGCTGAA CCGACTCCTG CAGGACAAGT
1551 GGGACAGATT CGTCAAGCGC ATCTTCTACT TCAACTTCCT GGTCTACTGC
1601 CTGTACATGA TCATCTTCAC CATGGCTGCC TACTACAGGC CCGTGGATGG
1651 CTTGCCTCCC TTTAAGATGG AAAAAATTGG AGACTATTTC CGAGTTACTG
1701 GAGAGATCCT GTCTGTGTTA GGAGGAGTCT ACTTCTTTTT CCGAGGGATT
1751 CAGTATTTCC TGCAGAGGCG GCCGTCGATG AAGACCCTGT TTGTGGACAG
1801 CTACAGTGAG ATGCTTTTCT TTCTGCAGTC ACTGTTCATG CTGGCCACCG
1851 TGGTGCTGTA CTTCAGCCAC CTCAAGGAGT ATGTGGCTTC CATGGTATTC
1901 TCCCTGGCCT TGGGCTGGAC CAACATGCTC TACTACACCC GCGGTTTCCA
1951 GCAGATGGGC ATCTATGCCG TCATGATAGA GAAGATGATC CTGAGAGACC
2001 TGTGCCGTTT CATGTTTGTC TACATCGTCT TCTTGTTCGG GTTTTCCACA
2051 GCGGTGGTGA CGCTGATTGA AGACGGGAAG AATGACTCCC TGCCGTCTGA
2101 GTCCACGTCG CACAGGTGGC GGGGGCCTGC CTGCAGGCCC CCCGATAGCT
2151 CCTACAACAG CCTGTACTCC ACCTGCCTGG AGCTGTTCAA GTTCACCATC
2201 GGCATGGGCG ACCTGGAGTT CACTGAGAAC TATGACTTCA AGGCTGTCTT
2251 CATCATCCTG CTGCTGGCCT ATGTAATTCT CACCTACATC CTCCTGCTCA
2301 ACATGCTCAT CGCCCTCATG GGTGAGACTG TCAACAAGAT CGCACAGGAG
2351 AGCAAGAACA TCTGGAAGCT GCAGAGAGCC ATCACCATCC TGGACACGGA
2401 GAAGAGCTTC CTTAAGTGCA TGAGGAAGGC CTTCCGCTCA GGCAAGCTGC
2451 TGCAGGTGGG GTACACACCT GATGGCAAGG ACGACTACCG GTGGTGCTTC
2501 AGGGTGGACG AGGTGAACTG GACCACCTGG AACACCAACG TGGGCATCAT
2551 CAACGAAGAC CCGGGCAACT GTGAGGGCGT CAAGCGCACC CTGAGCTTCT
2601 CCCTGCGGTC AAGCAGAGTT TCAGGCAGAC ACTGGAAGAA CTTTGCCCTG
2651 GTCCCCCTTT TAAGAGAGGC AAGTGCTCGA GATAGGCAGT CTGCTCAGCC
```

```
2701 CGAGGAAGTT TATCTGCGAC AGTTTTCAGG GTCTCTGAAG CCAGAGGACG
2751 CTGAGGTCTT CAAGAGTCCT GCCGCTTCCG GGGAGAAGTG AGGACGTCAC
2801 GCAGACAGCA CTGTCAACAC TGGGCCTTAG GAGACCCCGT TGCCACGGGG
2851 GGCTGCTGAG GGAACACCAG TGCTCTGTCA GCAGCCTGGC CTGGTCTGTG
2901 CCTGCCCAGC ATGTTCCCAA ATCTGTGCTG GACAAGCTGT GGGAAGCGTT
2951 CTTGGAAGCA TGGGGAGTGA TGTACATCCA ACCGTCACTG TCCCCAAGTG
3001 AATCTCCTAA CAGACTTTCA GGTTTTTACT CACTTTACTA AACAGTTTGG
3051 ATGGTCAGTC TCTACTGGGA CATGTTAGGC CCTTGTTTTC TTTGATTTTA
3101 TTCTTTTTTT TGAGACAGAA TTTCACTCTT CTCACCCAGG CTGGAATGCA
3151 GTGGCACAAT TTTGGCTCCC TGCAACCTCC GCCTCCTGGA TTCCAGCAAT
3201 TCTCCTGCCT CGGCTTCCCA AGTAGCTGGG ATTACAGGCA CGTGCCACCA
3251 TGTCTGGCTA ATTTTTTGTA TTTTTTTAAT AGATATGGGG TTTCGCCATG
3301 TTGGCCAGGC TGGTCTCGAA CTCCTGACCT CAGGTGATCC GCCCACCTCG
3351 GCCTCCCAAA GTGCTGGGAT TACAGGTGTG AGCCTCCACA CCTGGCTGTT
3401 TTCTTTGATT TTATTCTTTT TTTTTTTTCT GTGAGACAGA GTTTCACTCT
3451 TGTTGCCCAG GCTGGAGTGC AGTGGTGTGA TCTTGGCTCA CTGCAACCTC
3501 TGCCTCCCGG GTTCAAGCGA TTCTTCTGCT TCAGTCTCCC AAGTAGCTTG
3551 GATTACAGGT GAGCACTACC ACGCCCGGCT AATTTTTGTA TTTTTAATAG
3601 AGACGGGGTT TCACCATGTT GGCCAGGCTG GTCTCGAACT CTTGACCTCA
3651 GGTGATCTGC CCGCCTTGGC CTCCCAAAGT GCTGGGATTA CAGGTGTGAG
3701 CCGCTGCGCT CGGCCTTCTT TGATTTTATA TTATTAGGAG CAAAAGTAAA
3751 TGAAGCCCAG GAAAACACCT TTGGGAACAA ACTCTTCCTT TGATGGAAAA
3801 TGCAGAGGCC CTTCCTCTCT GTGCCGTGCT TGCTCCTCTT ACCTGCCCGG
3851 GTGGTTTGGG GGTGTTGGTG TTTCCTCCCT GGAGAAGATG GGGGAGGCTG
3901 TCCCACTCCC AGCTCTGGCA GAATCAAGCT GTTGCAGCAG TGCCTTCTTC
3951 ATCCTTCCTT ACGATCAATC ACAGTCTCCA GAAGATCAGC TCAATTGCTG
4001 TGCAGGTTAA AACTACAGAA CCACATCCCA AAGGTACCTG GTAAGAATGT
4051 TTGAAAGATC TTCCATTTCT AGGAACCCCA GTCCTGCTTC TCCGCAATGG
4101 CACATGCTTC CACTCCATCC ATACTGGCAT CCTCAAATAA ACAGATATGT
4151 ATACATATAA AAAAAAAAAA AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

99288727:
Recent advances in neuropharmacology of cutaneous nociceptors.

99231880:
A non-pungent triprenyl phenol of fungal origin, scutigeral, stimulates rat dorsal root ganglion
neurons via interaction at vanilloid receptors.

Peptide information for frame 2

ORF from 272 bp to 2788 bp; peptide length: 839
Category: strong similarity to known protein
Classification: Cell signaling/communication

```
  1 MKKWSSTDLG AAADPLQKDT CPDPLDGDPN SRPPPAKPQL STAKSRTRLF
 51 GKGDSEEAFP VDCPHEEGEL DSCPTITVSP VITIQRPGDG PTGARLLSQD
101 SVAASTEKTL RLYDRRSIFE AVAQNNCQDL ESLLLFLQKS KKHLTDNEFK
151 DPETGKTCLL KAMLNLHDGQ NTTIPLLLEI ARQTDSLKEL VNASYTDSYY
201 KGQTALHIAI ERRNMALVTL LVENGADVQA AAHGDFFKKT KGRPGFYFGE
251 LPLSLAACTN QLGIVKFLLQ NSWQTADISA RDSVGNTVLH ALVEVADNTA
301 DNTKFVTSMY NEILILGAKL HPTLKLEELT NKKGMTPLAL AAGTGKIGVL
351 AYILQREIQE PECRHLSRKF TEWAYGPVHS SLYDLSCIDT CEKNSVLEVI
401 AYSSSETPNR HDMLLVEPLN RLLQDKWDRF VKRIFYFNFL VYCLYMIIFT
451 MAAYYRPVDG LPPFKMEKIG DYFRVTGEIL SVLGGVYFFF RGIQYFLQRR
501 PSMKTLFVDS YSEMLFFLQS LFMLATVVLY FSHLKEYVAS MVFSLALGWT
551 NMLYYTRGFQ QMGIYAVMIE KMILRDLCRF MFVYIVFLFG FSTAVVTLIE
601 DGKNDSLPSE STSHRWRGPA CRPPDSSYNS LYSTCLELFK FTIGMGDLEF
651 TENYDFKAVF IILLLAYVIL TYILLLNMLI ALMGETVNKI AQESKNIWKL
701 QRAITILDTE KSFLKCMRKA FRSGKLLQVG YTPDGKDDYR WCFRVDEVNW
751 TTWNTNVGII NEDPGNCEGV KRTLSFSLRS SRVSGRHWKN FALVPLLREA
801 SARDRQSAQP EEVYLRQFSG SLKPEDAEVF KSPAASGEK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_20k2, frame 2

TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds., N = 1, Score = 3760, P = 0

TREMBLNEW:AB015231_1 product: "stretch-inhibitable nonselective channel (SIC)"; Rattus norvegicus mRNA for stretch-inhibitable nonselective channel (SIC), complete cds., N = 2, Score = 2090, P = 2e-219

```
>TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus
     norvegicus vanilloid receptor subtype 1 mRNA, complete cds.
            Length = 838

  HSPs:

 Score = 3760 (564.1 bits), Expect = 0.0e+00, P = 0.0e+00
 Identities = 721/839 (85%), Positives = 773/839 (92%)

Query:     1 MKKWSSTDLGAAADPLQKDTCPDPLDGDPNSRPPPAKPQLSTAKSRTRLFGKGDSEEAFP 60
             M++ +S D    +  P Q++C DP D DPN +PPP KP + T +SRTRLFGKGDSEEA P
Sbjct:     1 MEQRASLDSEESESPPQENSCLDPPDRDPNCKPPPVKPHIFTTRSRTRLFGKGDSEEASP 60

Query:    61 VDCPHEEGELDSCPTITVSPVITIQRPGDGPTGARLLSQDSVAASTEKTLRLYDRRSIFE 120
             +DCP+EEG L SCP ITVS V+TIQRPGDGP   R  SQDSV+A  EK  RLYDRRSIF+
Sbjct:    61 LDCPYEEGGLASCPIITVSSVLTIQRPGDGPASVRPSSQDSVSAG-EKPPRLYDRRSIFD 119

Query:   121 AVAQNNCQDLESLLLFLQKSKKHLTDNEFKDPETGKTCLLKAMLNLHDGQNTTIPLLLEI 180
             AVAQ+NCQ+LESLL FLQ+SKK LTD+EFKDPETGKTCLLKAMLNLH+GQN TI LLL++
Sbjct:   120 AVAQSNCQELESLLPFLQRSKKRLTDSEFKDPETGKTCLLKAMLNLHNGQNDTIALLLDV 179

Query:   181 ARQTDSLKELVNASYTDSYYKGQTALHIAIERRNMALVTLLVENGADVQAAAHGDFFKKT 240
             AR+TDSLK+ VNASYTDSYYKGQTALHIAIERRNM LVTLLVENGADVQAAA+GDFFKKT
Sbjct:   180 ARKTDSLKQFVNASYTDSYYKGQTALHIAIERRNMTLVTLLVENGADVQAAANGDFFKKT 239

Query:   241 KGRPGFYFGELPLSLAACTNQLGIVKFLLQNSWQTADISARDSVGNTVLHALVEVADNTA 300
             KGRPGFYFGELPLSLAACTNQL IVKFLLQNSWQ ADISARDSVGNTVLHALVEVADNT
Sbjct:   240 KGRPGFYFGELPLSLAACTNQLAIVKFLLQNSWQPADISARDSVGNTVLHALVEVADNTV 299

Query:   301 DNTKFVTSMYNEILILGAKLHPTLKLEELTNKKGMTPLALAAGTGKIGVLAYILQREIQE 360
             DNTKFVTSMYNEILILGAKLHPTLKLEE+TN+KG+TPLALAA +GKIGVLAYILQREI E
Sbjct:   300 DNTKFVTSMYNEILILGAKLHPTLKLEEITNRKGLTPLALAASSGKIGVLAYILQREIHE 359

Query:   361 PECRHLSRKFTEWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN 420
             PECRHLSRKFTEWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN
Sbjct:   360 PECRHLSRKFTEWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN 419

Query:   421 RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYYRPVDGLPPFKMEK-IGDYFRVTGEI 479
             RLLQDKWDRFVKRIFYFNF VYCLYMIIFT AAYYRPV+GLPP+K++  +GDYFRVTGEI
Sbjct:   420 RLLQDKWDRFVKRIFYFNFFVYCLYMIIFTAAAYYRPVEGLPPYKLKNTVGDYFRVTGEI 479

Query:   480 LSVLGGVYFFFRGIQYFLQRRPSMKTLFVDSYSEMLFFLQSLFMLATVVLYFSHLKEYVA 539
             LSV GGVYFFFRGIQYFLQRRPS+K+LFVDSYSE+LFF+QSLFML +VVLYFS  KEYVA
Sbjct:   480 LSVSGGVYFFFRGIQYFLQRRPSLKSLFVDSYSEILFFVQSLFMLVSVVLYFSQRKEYVA 539

Query:   540 SMVFSLALGWTNMLYYTRGFQQMGIYAVMIEKMILRDLCRFMFVYIVFLFGFSTAVVTLI 599
             SMVFSLA+GWTNMLYYTRGFQQMGIYAVMIEKMILRDLCRFMFVY+VFLFGFSTAVVTLI
Sbjct:   540 SMVFSLAMGWTNMLYYTRGFQQMGIYAVMIEKMILRDLCRFMFVYLVFLFGFSTAVVTLI 599

Query:   600 EDGKNDSLPSESTSHRWRGPACRPPDSSYNSLYSTCLELFKFTIGMGDLEFTENYDFKAV 659
             EDGKN+SLP EST H+ RG AC+P   +SYNSLYSTCLELFKFTIGMGDLEFTENYDFKAV
Sbjct:   600 EDGKNNSLPMESTPHKCRGSACKP-GNSYNSLYSTCLELFKFTIGMGDLEFTENYDFKAV 658

Query:   660 FIILLLAYVILTYILLLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSFLKCMRK 719
             FIILLLAYVILTYILLLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSFLKCMRK
Sbjct:   659 FIILLLAYVILTYILLLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSFLKCMRK 718

Query:   720 AFRSGKLLQVGYTPDGKDDYRWCFRVDEVNWTTWNTNVGIINEDPGNCEGVKRTLSFSLR 779
             AFRSGKLLQVG+TPDGKDDYRWCFRVDEVNWTTWNTNVGIINEDPGNCEGVKRTLSFSLR
Sbjct:   719 AFRSGKLLQVGFTPDGKDDYRWCFRVDEVNWTTWNTNVGIINEDPGNCEGVKRTLSFSLR 778

Query:   780 SSRVSGRHWKNFALVPLLREASARDRQSAQPEEVYLRQFSGSLKPEDAEVFKSPAASGEK 839
             S RVSGR+WKNFALVPLLR+AS RDR + Q EEV L+ ++GSLKPEDAEVFK      GEK
Sbjct:   779 SGRVSGRNWKNFALVPLLRDASTRDRHATQQEEVQLKHYTGSLKPEDAEVFKDSMVPGEK 838
```

Pedant information for DKFZphtes3_20k2, frame 2

Report for DKFZphtes3_20k2.2

```
[LENGTH]     839
[MW]         94950.75
[pI]         6.90
[HOMOL]      TREMBL:AF029310_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus
vanilloid receptor subtype 1 mRNA, complete cds. 0.0
[FUNCAT]     99 unclassified proteins     [S. cerevisiae, YIL112w] 4e-05
[PIRKW]      alternative splicing 3e-06
[PIRKW]      peripheral membrane protein 3e-06
[SUPFAM]     ankyrin repeat homology 3e-06
[SUPFAM]     unassigned ankyrin repeat proteins 3e-06
[PFAM]       Ank repeat
[KW]         TRANSMEMBRANE  4
```

```
SEQ   MKKWSSTDLGAAADPLQKDTCPDPLDGDPNSRPPPAKPQLSTAKSRTRLFGKGDSEEAFP
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM   ............................................................

SEQ   VDCPHEEGELDSCPTITVSPVITIQRPGDGPTGARLLSQDSVAASTEKTLRLYDRRSIFE
PRD   ccccccccccccccccceeeeeeecccccccceeeccccccccccccchhhhhhhhhhhh
MEM   ............................................................

SEQ   AVAQNNCQDLESLLLFLQKSKKHLTDNEFKDPETGKTCLLKAMLNLHDGQNTTIPLLLEI
PRD   hhhhcchhhhhhhhhhhhhhcccccccccccccccchhhhhhhhhhccccccchhhhhhh
MEM   ............................................................

SEQ   ARQTDSLKELVNASYTDSYYKGQTALHIAIERRNMALVTLLVENGADVQAAAHGDFFKKT
PRD   hhhccccccccccccccccccchhhhhhhhhchhhhhhhhhccceeeccccccccccccc
MEM   ............................................................

SEQ   KGRPGFYFGELPLSLAACTNQLGIVKFLLQNSWQTADISARDSVGNTVLHALVEVADNTA
PRD   ccccceeeccccchhhhhhcchhhhhhhhhcccccccccccccccchhhhhhhhhhcccc
MEM   ............................................................

SEQ   DNTKFVTSMYNEILILGAKLHPTLKLEELTNKKGMTPLALAAGTGKIGVLAYILQREIQE
PRD   chhhhhhhhhhhhhhhhccccccceeeeeecccccccchhhhhhhcchhhhhhhhhhhhc
MEM   ............................................................

SEQ   PECRHLSRKFTEWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN
PRD   cccchhhhheeeccceeeeeeeecccccccccccccceeeeeccccccccccceeeeehhhh
MEM   ............................................................

SEQ   RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYYRPVDGLPPFKMEKIGDYFRVTGEIL
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccchhhhhhhc
MEM   .............MMMMMMMMMMMMMMMMM..............................

SEQ   SVLGGVYFFFRGIQYFLQRRPSMKTLFVDSYSEMLFFLQSLFMLATVVLYFSHLKEYVAS
PRD   cccceeeeeecchhhhhhhhheeeeeeccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM   ..................................MMMMMMMMMMMMMMMMM.........

SEQ   MVFSLALGWTNMLYYTRGFQQMGIYAVMIEKMILRDLCRFMFVYIVFLFGFSTAVVTLIE
PRD   hhhhhhhhhhhhheeecccccccccchhhhhhhhhhhhhhhhhhheeeccccccceeeeeec
MEM   ..........................................MMMMMMMMMMMMMMMMM.

SEQ   DGKNDSLPSESTSHRWRGPACRPPDSSYNSLYSTCLELFKFTIGMGDLEFTENYDFKAVF
PRD   ccccccccccccccccccccccccccccccccchhhhhhhhhhhhcccchhhhhhhhhhh
MEM   ..........................................................MM

SEQ   IILLLAYVILTYILLLNMLIALMGETVNKIAQESKNIWKLQRAITILDTEKSFLKCMRKA
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM   MMMMMMMMMMMMMMM.............................................

SEQ   FRSGKLLQVGYTPDGKDDYRWCFRVDEVNWTTWNTNVGIINEDPGNCEGVKRTLSFSLRS
PRD   hhcceeeeeeecccccccccceeeeeeeeccccccccccceeeeeccccccccceeeeeeeeec
MEM   ............................................................

SEQ   SRVSGRHWKNFALVPLLREASARDRQSAQPEEVYLRQFSGSLKPEDAEVFKSPAASGEK
PRD   cccccccccchhhhhhhhhhhhhhhcccccceeeeeccccccccccccceeeecccccccc
MEM   ...........................................................
```

(No Prosite data available for DKFZphtes3_20k2.2)

Pfam for DKFZphtes3_20k2.2

```
HMM_NAME        Ank repeat

HMM                 *GyTPLHIAARyNNvEMVrlLLQHGADIN*
                     G+T+LHIA +++N+ +V LL+++GAD+
Query           202  GQTALHIAIERRNMALVTLLVENGADVQ    229
```

DKFZphtes3_2013
----

group: transmembrane protein

DKFZphtes3_2013 encodes a novel 595 amino acid protein with partial similarity to the IL-17 receptor.

The novel protein contains one transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to IL-17 receptor

Sequenced by MediGenomix

Locus: unknown

Insert length: 2406 bp
Poly A stretch at pos. 2345, no polyadenylation signal found

```
   1 GCCTCAGGTG TTCCTGCGTT GTTTGTCAGT GGAGAGCAGG GAGTGGGGCC
  51 AGCCAGCAGA AACAGTGGGC TGTACAACAT CACCTTCAAA TATGACAATT
 101 GTACCACCTA CTTGAATCCA GTGGGGAAGC ATGTGATTGC TGACGCCCAG
 151 AATATCACCA TCAGCCAGTA TGCTTGCCAT GACCAAGTGG CAGTCACCAT
 201 TCTTTGGTCC CCAGGGGCCC TCGGCATCGA ATTCCTGAAA GGATTTCGGG
 251 TAATACTGGA GGAGCTGAAG TCGGAGGGAA GACAGTGCCA ACAACTGATT
 301 CTAAAGGATC CGAAGCAGCT CAACAGTAGC TTCAAAAGAA CTGGAATGGA
 351 ATCTCAACCT TTCCTGAATA TGAAATTTGA AACGGATTAT TTCGTAAAGG
 401 TTGTCCCTTT TCCTTCCATT AAAAACGAAA GCAATTACCA CCCTTTCTTC
 451 TTTAGAACCC GAGCCTGTGA CCTGTTGTTA CAGCCGGACA ATCTAGCTTG
 501 TAAACCCTTC TGGAAGCCTC GGAACCTGAA CATCAGCCAG CATGGCTCGG
 551 ACATGCAGGT GTCCTTCGAC CACGCACCGC ACAACTTCGG CTTCCGTTTC
 601 TTCTATCTTC ACTACAAGCT CAAGCACGAA GGACCTTTCA AGCGAAAGAC
 651 CTGTAAGCAG GAGCAAACTA CAGAGATGAC CAGCTGCCTC CTTCAAAATG
 701 TTTCTCCAGG GGATTATATA ATTGAGCTGG TGGATGACAC TAACACAACA
 751 AGAAAAGTGA TGCATTATGC CTTAAAGCCA GTGCACTCCC CGTGGGCCGG
 801 GCCCATCAGA GCCGTGGCCA TCACAGTGCC ACTGGTAGTC ATATCGGCAT
 851 TCGCGACGCT CTTCACTGTG ATGTGCCGCA AGAAGCAACA AGAAAATATA
 901 TATTCACATT TAGATGAAGA GAGCTCTGAG TCTTCCACAT ACACTGCAGC
 951 ACTCCCAAGA GAGAGGCTCC GGCCGCGGCC GAAGGTCTTT CTCTGCTATT
1001 CCAGTAAAGA TGGCCAGAAT CACATGAATG TCGTCCAGTG TTTCGCCTAC
1051 TTCCTCCAGG ACTTCTGTGG CTGTGAGGTG GCTCTGGACC TGTGGGAAGA
1101 CTTCAGCCTC TGTAGAGAAG GGCAGAGAGA ATGGGTCATC CAGAAGATCC
1151 ACGAGTCCCA GTTCATCATT GTGGTTTGTT CCAAAGGTAT GAAGTACTTT
1201 GTGGACAAGA AGAACTACAA ACACAAAGGA GGTGGCCGAG GCTCGGGGAA
1251 AGGAGAGCTC TTCCTGGTGG CGGTGTCAGC CATTGCCGAA AAGCTCCGCC
1301 AGGCCAAGCA GAGTTCGTCC GCGGCGCTCA GCAAGTTTAT CGCCGTCTAC
1351 TTTGATTATT CCTGCGAGGG AGACGTCCCC GGTATCCTAG ACCTGAGTAC
1401 CAAGTACAGA CTCATGGACA ATCTTCCTCA GCTCTGTTCC CACCTGCACT
1451 CCCGAGACCA CGGCCTCCAG GAGCCGGGGC AGCACACGCG ACAGGGCAGC
1501 AGAAGGAACT ACTTCCGGAG CAAGTCAGGC CGGTCCCTAT ACGTCGCCAT
1551 TTGCAACATG CACCAGTTTA TTGACGAGGA GCCCGACTGG TTCGAAAAGC
1601 AGTTCGTTCC CTTCCATCCT CCTCCACTGC GCTACCGGGA GCCAGTCTTG
1651 GAGAAATTTG ATTCGGGCTT GGTTTTAAAT GATGTCATGT GCAAACCAGG
1701 GCCTGAGAGT GACTTCTGCC TAAAGGTAGA GGCGGCTGTT CTTGGGGCAA
1751 CCGGACCAGC CGACTCCCAG CACGAGAGTC AGCATGGGGG CCTGGACCAA
1801 GACGGGGAGG CCCGGCCTGC CCTTGACGGT AGCGCCGCCC TGCAACCCCT
1851 GCTGCACACG GTGAAAGCCG GCAGCCCCTC GGACATGCCG CGGGACTCAG
1901 GCATCTATGA CTCGTCTGTG CCCTCATCCG AGCTGTCTCT GCCACTGATG
1951 GAAGGACTCT CGACGGACCA GACAGAAACG TCTTCCCTGA CGGAGAGCGT
2001 GTCCTCCTCT TCAGGCCTGG GTGAGGAGGA ACCTCCTGCC CTTCCTTCCA
2051 AGCTCCTCTC TTCTGGGTCA TGCAAAGCAG ATCTTGGTTG CCGCAGCTAC
2101 ACTGATGAAC TCCACGCGGT CGCCCCTTTG TAACAAAACG AAAGAGTCTA
2151 AGCATTGCCA CTTTAGCTGC TGCCTCCCTC TGATTCCCCA GCTCATCTCC
2201 CTGGTTGCAT GGCCCACTTG GAGCTGAGGT CTCATACAAG GATATTTGGA
2251 GTGAAATGCT GGCCAGTACT TGTTCTCCCT TGCCCCAACC CTTTACCGGA
2301 TATCTTGACA AACTCTCCAA TTTTCTAAAA TGATATGGAG CTCTGAAAAA
2351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2401 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 346 bp to 2130 bp; peptide length: 595
Category: similarity to known protein
Classification: unclassified

```
  1 MESQPFLNMK FETDYFVKVV PFPSIKNESN YHPFFFRTRA CDLLLQPDNL
 51 ACKPFWKPRN LNISQHGSDM QVSFDHAPHN FGFRFFYLHY KLKHEGPFKR
101 KTCKQEQTTE MTSCLLQNVS PGDYIIELVD DTNTTRKVMH YALKPVHSPW
151 AGPIRAVAIT VPLVVISAFA TLFTVMCRKK QQENIYSHLD EESSESSTYT
201 AALPRERLRP RPKVFLCYSS KDGQNHMNVV QCFAYFLQDF CGCEVALDLW
251 EDFSLCREGQ REWVIQKIHE SQFIIVVCSK GMKYFVDKKN YKHKGGGRGS
301 GKGELFLVAV SAIAEKLRQA KQSSSAALSK FIAVYFDYSC EGDVPGILDL
351 STKYRLMDNL PQLCSHLHSR DHGLQEPGQH TRQGSRRNYF RSKSGRSLYV
401 AICNMHQFID EEPDWFEKQF VPFHPPPLRY REPVLEKFDS GLVLNDVMCK
451 PGPESDFCLK VEAAVLGATG PADSQHESQH GGLDQDGEAR PALDGSAALQ
501 PLLHTVKAGS PSDMPRDSGI YDSSVPSSEL SLPLMEGLST DQTETSSLTE
551 SVSSSSGLGE EEPPALPSKL LSSGSCKADL GCRSYTDELH AVAPL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2013, frame 1

TREMBL:U58917_1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor mRNA, complete cds., N = 1, Score = 215, P = 4.7e-14

TREMBL:MM31993_1 product: "interleukin 17 receptor"; Mus musculus interleukin 17 receptor mRNA, complete cds., N = 2, Score = 152, P = 1.1e-13

>TREMBL:U58917_1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor mRNA, complete cds.
Length = 866

HSPs:

Score = 215 (32.3 bits), Expect = 4.7e-14, P = 4.7e-14
Identities = 85/284 (29%), Positives = 131/284 (46%)

```
Query:   213 KVFLCYSSKDGQNHMNVVQCFAYFLQDFCGCEVALDLWEDFSLCREGQREWV-IQK---I 268
             KV++ YS+ D    +++VV  FA FL   CG EVALDL E+ ++   G   WV  QK   +
Sbjct:   379 KVWIIYSA-DHPLYVDVVLKFAQFLLTACGTEVALDLLEEQAISEAGVMTWVGRQKQEMV 437

Query:   269 HESQFIIVVCSKGMKY----FVDKKNYXXXXXXXXXXXXXELFLVAVSAIAEXXXXXXXXX 324
              +  IIV+CS+G +        + +                  +LF  A++ I
Sbjct:   438 ESNSKIIVLCSRGTRAKWQALLGRGAPVRLRCDHGKPVGDLFTAAMNMILPDFKRPACFG 497

Query:   325 XXXXXXFIAVYF-DYSCEGDVPGILDLSTKYRLMDNLPQLCSHLHSRDHGLQEPGQHTRQ 383
                   ++  YF + SC+GDVP +    + +Y LMD   ++  +   +D  + +PG+  R
Sbjct:   498 T-----YVVCYFSEVSCDGDVPDLFGAAPRYPLMDRFEEV--YFRIQDLEMFQPGRMHRV 550

Query:   384 G--SRRNYFRSKSGRSLYVAICNMHQFIDEEPDWFEKQFV----PFHPPPLR---YREPV 434
             G  S  NY RS  GR L  A+     +     PDWFE + +        P L     + EP+
Sbjct:   551 GELSGDNYLRSPGGRQLRAALDRFRDWQVRCPDWFECENLYSADDQDAPSLDEEVFEEPL 610

Query:   435 LEKFDSGLVLNDVMCKPGPESDFCLKVEAAVLGATGPADSQHESQHGGLDQDGEARP 491
             L     +G+V     + +  P S  CL ++  V G  G A        H  L  G+ P
Sbjct:   611 LPP-GTGIVKRAPLVRE-PGSQACLAIDPLV-GEEGGAAVAKLEPH--LQPRGQPAP 662
```

Pedant information for DKFZphtes3_2013, frame 1

Report for DKFZphtes3_2013.1

```
[LENGTH]        595
[MW]            66847.05
[pI]            6.27
[HOMOL]         TREMBL:MM31993_1 product: "interleukin 17 receptor"; Mus musculus interleukin
17 receptor mRNA, complete cds. 2e-14
[BLOCKS]        BL00740A MAM domain proteins
[BLOCKS]        BL01224B N-acetyl-gamma-glutamyl-phosphate reductase proteins
[KW]            TRANSMEMBRANE  1
[KW]            LOW_COMPLEXITY    13.61 %

SEQ    MESQPFLNMKFETDYFVKVVPFPSIKNESNYHPFFFRTRACDLLLQPDNLACKPFWKPRN
SEG    ............................................................
PRD    cccccccccccccceeeeecccccccccccccceeeeeeceeeeeeecccccccccccccc
MEM    ............................................................

SEQ    LNISQHGSDMQVSFDHAPHNFGFRFFYLHYKLKHEGPFKRKTCKQEQTTEMTSCLLQNVS
SEG    ............................................................
PRD    eeeecccccceeeecccccccceeeeeehhhhhcccchhhhhhhhhhhhhhhhhhhcccc
MEM    ............................................................

SEQ    PGDYIIELVDDTNTTRKVMHYALKPVHSPWAGPIRAVAITVPLVVISAFATLFTVMCRKK
SEG    ............................................................
PRD    ccceeeeeeccccccccccccccccccccccccccceeeecchhhhhhhhhhhhhhhhhh
MEM    ....................................MMMMMMMMMMMMMMMMM.......

SEQ    QQENIYSHLDEESSESSTYTAALPRERLRPRPKVFLCYSSKDGQNHMNVVQCFAYFLQDF
SEG    ..........xxxxxxx....xxxxxxxxxx.............................
PRD    hhhhhhhhccccccceeeecccccccccccccceeeeeeecccccchhhhhhhhhhhhhhhc
MEM    ............................................................

SEQ    CGCEVALDLWEDFSLCREGQREWVIQKIHESQFIIVVCSKGMKYFVDKKNYKHKGGGRGS
SEG    ...................................................xxxxxxxxx
PRD    ccchhhhhhhcccccccchhhhhhhhhheeeeeeeeccceeeeeeccccccccccccccc
MEM    ............................................................

SEQ    GKGELFLVAVSAIAEKLRQAKQSSSAALSKFIAVYFDYSCEGDVPGILDLSTKYRLMDNL
SEG    xxx............xxxxxxxxxxxxxxx..............................
PRD    ccceeeehhhhhhhhhhhhhhcchhhhhhheeeecccccccccccccccccchhhhhhccc
MEM    ............................................................

SEQ    PQLCSHLHSRDHGLQEPGQHTRQGSRRNYFRSKSGRSLYVAICNMHQFIDEEPDWFEKQF
SEG    ............................................................
PRD    cchhhhhcccccccccccccccccceeeecccccccceeeeeeeceeeecccccceeeeee
MEM    ............................................................

SEQ    VPFHPPPLRYREPVLEKFDSGLVLNDVMCKPGPESDFCLKVEAAVLGATGPADSQHESQH
SEG    ............................................................
PRD    eeccccccccccceeeeeccccceeeeeccccccccccchhhhhhhhhcccccccccccc
MEM    ............................................................

SEQ    GGLDQDGEARPALDGSAALQPLLHTVKAGSPSDMPRDSGIYDSSVPSSELSLPLMEGLST
SEG    ..........................................xxxxxxxxxxxxxxxxx.
PRD    cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchh
MEM    ............................................................

SEQ    DQTETSSLTESVSSSSGLGEEEPPALPSKLLSSGSCKADLGCRSYTDELHAVAPL
SEG    ..xxxxxxxxxxxxxxxxxx...................................
PRD    hhhhhhhheeccccccccccccccccceeecccccceeecccccccccceeeeccc
MEM    .......................................................
```

(No Prosite data available for DKFZphtes3_2013.1)

(No Pfam data available for DKFZphtes3_2013.1)

DKFZphtes3_20m18
----

group: nucleic acid management

DKFZphtes3_20m18 encodes a novel 132 amino acid protein with similarity to the S. cerevisiae mitochondrial carrier protein RIM2.

The novel protein contains a leucine zipper and a Prosite mitochondrial energy transfer proteins signature. It is member of a family of substrate carrier proteins which are found in the inner mitochondrial membrane and are involved in energy transfer. The RIM2/MRS12 gene encodes a predicted protein of 377 amino acids that is essential for mitochondrial DNA metabolism and proper cell growth. Inactivation of this gene causes the total loss of mitochondrial DNA and, compared to wild-type rhoo controls, a slow-growth phenotype on media containing glucose. The novel protein seems to be the human orthologue of this protein.

The new protein can find application in modulation of mitochondrial DNA replication and maintenance.

similarity to carrier protein RIM2

Sequenced by MediGenomix

Locus: unknown

Insert length: 3572 bp
Poly A stretch at pos. 3530, polyadenylation signal at pos. 3510

```
   1 GCCGCGGGGA GGGCTGTGCC GGTTGCTTTC TGCAGCCGCA TCTCGGCCAG
  51 CTCTCCTCGC CGTCCCCGGG GCGCTGTGCG TCTCCAGTCC GGGACCGAAG
 101 CCGCCTGCCG TAGCGGGCGG CCAGATCCGC GTCCCGCCTC AGCGGCCGGA
 151 GGACATGCGG GAGAGAGAAT GAGCCAGAGG GACACGCTGG TGCATCTGTT
 201 TGCCGGAGGA TGTGGTGGTA CAGTGGGAGC TATTCTGACA TGTCCACTGG
 251 AAGTTGTAAA AACACGACTG CAGTCATCTT CTGTGACGCT TTATATTTCT
 301 GAAGTTCAGC TGAACACCAT GGCTGGAGCC AGTGTCAACC GAGTAGTGTC
 351 TCCCGGACCT CTTCATTGCC TAAAGGTGAT CTTGGAAAAA GAAGGGCCTC
 401 GTTCCTTGTT TAGAGGACTA GGCCCCAATT TAGTGGGGGT AGCCCCTTCC
 451 AGAGCAATAT ACTTTGCTGC TTATTCAAAC TGCAAGGAAA AGTTGAATGA
 501 TGTATTTGAT CCTGATTCTA CCCAAGTACA TATGATTTCA GCTGCAATGG
 551 CAGGTATGAA TGTATAATAT TAAAAAAAAA AAAAACTTTC TGAAACCTAG
 601 AGGCTTAATA TTGAATTATA AGTTTGTAGT GAAAAGTTGA TGATTAATGT
 651 GCTTTTCATT GATTAGATGA TTTTTACGTT TATCGATATA AACCAAATTA
 701 GGTATATGTA AAATCTGTCA TCAGTTGACA TTTTTGTAGT CAGGAGTTTA
 751 CATGCTAGGG TACAAGTAAT ATATTTATAT TGCCTTGTGT AGTCCACTGA
 801 ATGTTTAGTG ATCATTGTTA ACAGTTTTAA GAATCCAACC ATAATTACAC
 851 TATAAATAAG TTATGGAGCT GTAATTTACT CTTCTCTCCT CAATTTCTGT
 901 TAGTGCCTTT TCCCTTTTTG CTGCATGTTT TGGCTTCTGT CTGAAATGTG
 951 TCGGCAATTC TTGGTAAAGT ATTCATTTTG TCCTGTGCTC AAATGCTGAA
1001 ATTTTTGTGA GTGATGTATT ATTATTGACA ATTCAGTTAC TATGTGTATT
1051 TTTTAAAATT GTTTATTATT CTACATAATT CACACTAGAC AGCACCTGAA
1101 ATTTAGACAC TGGCTATGTG TACATGCTTA CTATAGAAAT GTTTCCAGGA
1151 ACTCTCTGTT TCTGTCATCA CTGATAAGTA TATATGATTC TGAATTAAAA
1201 TAACTAGTTT TAGGTCTTTA CCCTGCCATA AAGATAAACA GTTGGTTTGA
1251 CCAATCTGGT TCTGGAATCA TTTGCTGCTA TGCATGTTAG ACAAAGCCAC
1301 GAACTTTGAT TTTCCATTGA AAATTCTCCC TAATATCTGA GATTTATTGT
1351 ATATTTACTC ATATCTCACA TTTTCAAATT ATGCTGTAAC TTTATAAACT
1401 GTAGCTGCTT TCATCAGCTA TTGATCAATA AATTGAATGT CAATTATGTG
1451 CTTAATAATG AGTGCCTTAA ACTGTTAAAC ACTTTTGGTT TAGAAATAAA
1501 GTGAATCAAT TTGACCTATA TACTTCATGA AGTAAGTAAG TTTGAAATAC
1551 AAATTTCTGA AAGGTCAATA GCCCTTATCG TATTACAAAT TGTTTTTAAG
1601 GCTTTTTGTA TTTATTAATT GTCAGTTGAT TCACTGAAGC TTTAAAACTG
1651 GAAGGGACAA TCCAAAGGTC AAAAGAGTGA AATACAATCA TTTACCAATA
1701 AGGAAACCTT GGGCAAATTA TGTAATTTAT GTGAACCTCT CTTAGCTTAC
1751 CCATGGAATG AGTCAAGTGG TCTACATAGA TTTGGATTTT GAGAATTAGT
1801 TCTTTCATTT AGTGTTATAG AGATTATCTT GTTACAACTA GAATTATTTT
1851 TAATGTAATT TTTACAGATG TTGAATATTA GTAGATAGGA TTTTTCCCCT
1901 ACGAATTTGG ATGTAAGGTA AAGGTTGGTG GCCAGTGACA AACCTTATAA
1951 CCACTTTATC AGGTTCTTTA AAAATATATT TGTGAATTAC CAGTGATTAT
2001 GTTTTTGGCT TATAACCTCA GATAATTATA AAGAAATGTT AATCTTATTT
2051 GAAAGAATTG GAATCTAGAA AGTTAGATGA GCAGTCATTT TATATTGATA
2101 TTTGTTATAT CAGTATAGCA AATGCAGAGG TTCAGAATAT CTTTATTTCC
2151 ACTGGAACAT CTTATTTCAT TAGAGTATCT CATCAGAATT TATTACTGTA
2201 TTTGTATCAC ATTGCAAAGA ATTTCAGTAG AATTGTCAGT TTGCACTTTT
2251 TTCTCAAATG TGTACAAATG TTAACATATA GTTCATTTTT ATCTGTACAT
2301 TGATGCCATT TCCCAACTTG AATTCCTCAA GTTTTGGTAA ACTTACAATC
2351 TCATACTTGT TCAGAGGTTA TTGCACTGTA CACTTACTGT GTAGAAAATA
2401 CTGTTTGAAT TTGTTTGCAG TTACATTGTT CTGAGAACTG TGCTCTCAGA
2451 GCTTCTGTGC ACTATTCATG AGCATTAACA CTTAGCCTTG CAGTTTTATA
```

```
2501 CATAACTATA TGGTTAGTAA AACTGAATGG TCCAATGCAG ACTCATTAAA
2551 GTAGGCTTTT GCCCCCTTTG TTCTTGAAAT AATCTAGACC AGATTACTCG
2601 GGGTTTTTTT TAGGATTATT TTTATAGGTC TAAATATGAA TGATTTGGGG
2651 GTATGAAGTA CTTAAAGATA GTTCTGTGAA AAATCATTTT CAGCTGTCTA
2701 TTCAAGGGAA AAAATGCTAA CCTTGTCACT TTACTACACA AAACCACACT
2751 AAAATAAACC ATTAATGATA CTGCCTGCAA GATTTTAACA CACCAGATAG
2801 CACACACATT AAGGATTTAT AAGGCACTGT ACGTAATTTT TATTCCAAGT
2851 GACCTCTCAA TTCATTTTCA TTTTGCATTT TATCCATATG AACTCATGTT
2901 TAATTTAGAT AATAAAAATT TATTTTATTA AAAGGACAGT TTATTTAAAG
2951 TGGGTCTTTT TATTTGTTGT AGTGCATACT ATAAGAATTT GTAAGCCTCT
3001 AAAGTTGAGC TATAAATTTT CATGCATTAA AAATTTGTTT CAGTTGTGAG
3051 GATATTTAAT CAGATTAAAT AATGTTGACT CTTAATATTT TGCCTGCCTT
3101 TTTTTTCTCC TACACATGAC CTTTGACAGA CTAAGTATAT CTCAGCTATT
3151 GAGGGTATCT GTTTTGTTGC CTGTATATTT TGTTTAAATT AACTTGTATA
3201 TTCCTTTGTA TACACCTAGG CACAGATGTA TGCAAAAAAA ATTTGTTAAA
3251 TTACTTCTTT CTTTATACTA ATTCTCAATT TTTAAAAGAT TTTATCTGGC
3301 ATGTATATAC TTTTATATAG AACATTATAA ATGTAAAGGA AATGAATTCT
3351 AATTTTAATT GGATTATGTA TTCATACAGT TATTCTCAAT TTTTAAAATA
3401 CTAATAATGT AATCATTGAA TGTTTCCTAC ATACGTAGTG GGTTTTATTT
3451 GCTCACAGCA TACAGTTATT TTTCAATTTA TGTTTTTCTA TTAGACTTAA
3501 ATTTCATTAT AATAAAGGCT TTTACTCATT AAATACAAAA AAAAAAAAAA
3551 AAAAAAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

95198680:
Overexpression of a novel member of the mitochondrial carrier family rescues defects in both DNA and RNA metabolism in yeast mitochondria.

Peptide information for frame 1

ORF from 169 bp to 564 bp; peptide length: 132
Category: similarity to known protein
Classification: Intacellular transport and traffic
Prosite motifs: LEUCINE_ZIPPER (27-49)
MITOCH_CARRIER (26-36)

```
  1 MSQRDTLVHL FAGGCGGTVG AILTCPLEVV KTRLQSSSVT LYISEVQLNT
 51 MAGASVNRVV SPGPLHCLKV ILEKEGPRSL FRGLGPNLVG VAPSRAIYFA
101 AYSNCKEKLN DVFDPDSTQV HMISAAMAGM NV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_20m18, frame 1

PIR:S44092 probable carrier protein c2 - Caenorhabditis elegans, N = 2,
Score = 147, P = 1.5e-19

PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast
(Saccharomyces cerevisiae), N = 1, Score = 230, P = 6.2e-19

>PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast
(Saccharomyces cerevisiae)
Length = 377

HSPs:

Score = 230 (34.5 bits), Expect = 6.2e-19, P = 6.2e-19
Identities = 55/133 (41%), Positives = 80/133 (60%)

```
Query:    8 VHLFAGGCGGTVGAILTCPLEVVKTRLQSSS-VTLYISEVQLNTMAGA----SVNRVVSP 62
            VH  AGG GG  GA++TCP ++VKTRLQS   +  Y S+  +N   G+     S+N V+
Sbjct:   54 VHFVAGGIGGMAGAVVTCPFDLVKTRLQSDIFLKAYKSQA-VNISKGSTRPKSINYVIQA 112
```

```
Query:    63 GP-----LHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAYSNCKEKLNDVFD--P 115
             G      L  +   + ++EG RSLF+GLGPNLVGV P+R+I F  Y    K+      F+
Sbjct:   113 GTHFKETLGIIGNVYKQEGFRSLFKGLGPNLVGVIPARSINFFTYGTTKDMYAKAFNNGQ 172

Query:   116 DSTQVHMISAAMAG 129
             ++  +H+++AA AG
Sbjct:   173 ETPMIHLMAAATAG 186

 Score = 77 (11.6 bits), Expect = 1.1e+00, P = 6.8e-01
 Identities = 25/88 (28%), Positives = 39/88 (44%)

Query:     3 QRDTLVHLFAGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVVSP 62
             Q    ++HL A    G   A  T P+ ++KTR              VQL+     SV +  +
Sbjct:   172 QETPMIHLMAAATAGWATATATNPIWLIKTR------------VQLDKAGKTSVRQYKNS 219

Query:    63 GPLHCLKVILEKEGPRSLFRGLGPNLVG 90
               CLK ++  EG   L++GL  + +G
Sbjct:   220 WD--CLKSVIRNEGFTGLYKGLSASYLG 245

 Score = 71 (10.7 bits), Expect = 6.6e+00, P = 1.0e+00
 Identities = 28/91 (30%), Positives = 45/91 (49%)

Query:    12 AGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVVSPGPLHCLKVI 71
             + G    V +I  T P EVV+TRL+ +                + N   G    R + G +  KVI
Sbjct:   294 SAGLAKFVASIATYPHEVVRTRLRQTP--------KEN---G---KRKYT-GLVQSFKVI 338

Query:    72 LEKEGPRSLFRGLGPNLVGVAPSRAIYFAAY 102
             +++EG  S++ GL P+L+   P+  I F  +
Sbjct:   339 IKEEGLFSMYSGLTPHLMRTVPNSIIMFGTW 369
```

Pedant information for DKFZphtes3_20m18, frame 1

Report for DKFZphtes3_20m18.1

```
[LENGTH]     132
[MW]         13993.36
[pI]         8.42
[HOMOL]      PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast (Saccharomyces cerevisiae) 7e-19
[FUNCAT]     07.16 purine and pyrimidine transporters    [S. cerevisiae, YBR192w] 3e-20
[FUNCAT]     08.04 mitochondrial transport    [S. cerevisiae, YBR192w] 3e-20
[FUNCAT]     30.16 mitochondrial organization    [S. cerevisiae, YBR192w] 3e-20
[FUNCAT]     02.13 respiration    [S. cerevisiae, YBR192w] 3e-20
[FUNCAT]     01.05.07 carbohydrate transport    [S. cerevisiae, YPR021c] 3e-10
[FUNCAT]     07.07 sugar and carbohydrate transporters    [S. cerevisiae, YPR021c] 3e-10
[FUNCAT]     07.99 other transport facilitators    [S. cerevisiae, YEL006w] 1e-09
[FUNCAT]     01.07.10 transport of vitamins, cofactors, and prosthetic groups    [S. cerevisiae, YIL006w] 3e-09
[FUNCAT]     07.04.07 anion transporters (cl, so4, po4, etc.)    [S. cerevisiae, YKL120w] 2e-08
[FUNCAT]     01.03.19 nucleotide transport    [S. cerevisiae, YPR011c] 3e-08
[FUNCAT]     04.05.03 mrna processing (splicing)    [S. cerevisiae, YKR052c] 4e-08
[FUNCAT]     01.05.04 regulation of carbohydrate utilization    [S. cerevisiae, YJR095w] 2e-07
[FUNCAT]     01.01.07 amino-acid transport    [S. cerevisiae, YOR130c] 5e-05
[FUNCAT]     07.10 amino-acid transporters    [S. cerevisiae, YOR130c] 5e-05
[FUNCAT]     01.04.07 phosphate transport  [S. cerevisiae, YJR077c] 7e-05
[FUNCAT]     13.04 homeostasis of other ions    [S. cerevisiae, YJR077c] 7e-05
[BLOCKS]     BL00215B Mitochondrial energy transfer proteins
[BLOCKS]     BL00215A Mitochondrial energy transfer proteins
[PIRKW]      duplication 6e-09
[PIRKW]      transmembrane protein 6e-09
[PIRKW]      mitochondrial inner membrane 4e-07
[PIRKW]      transport protein 5e-06
[PIRKW]      mitochondrion 7e-08
[PIRKW]      chloroplast 3e-08
[SUPFAM]     Btl protein 3e-08
[SUPFAM]     ADP,ATP carrier protein repeat homology 4e-09
[SUPFAM]     Caenorhabditis probable carrier protein c2 4e-09
[SUPFAM]     probable carrier protein YPR021c 6e-09
[PROSITE]    LEUCINE_ZIPPER 1
[PROSITE]    MITOCH_CARRIER 1
[PFAM]       Mitochondrial carrier proteins
[KW]         Alpha_Beta

SEQ     MSQRDTLVHLFAGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVV
```

```
PRD    cccccceeeecccccccceeeeeecchhhhhhhhhhcccccccccccccccccccccc

SEQ    SPGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAYSNCKEKLNDVFDPDSTQV
PRD    ccccchhhhhhhhhhccccceeeeccccceeeeccccceeeeeehhhhhhhhhccccccccc

SEQ    HMISAAMAGMNV
PRD    chhhhhhhcccc
```

Prosite for DKFZphtes3_20m18.1

PS00029	27->49	LEUCINE_ZIPPER	PDOC00029
PS00215	26->36	MITOCH_CARRIER	PDOC00189

Pfam for DKFZphtes3_20m18.1

```
HMM_NAME     Mitochondrial carrier proteins

HMM             *pFwkdFLAGGIAGmMeHTvMFPIDtIKTRMQlQgEMpM..ahpR.....
                 +++++++AGG +G + +++++P++++KTR+Q++  ++    + ++
Query         5  DTLVHLFAGGCGGTVGAILTCPLEVVKTRLQSS-SVTLYISEVQLNTMA    52

HMM             ........YkGMIdCFRwIwkNEGWRGLWRGLgANvIRYIPqWaIRFGFY
                             G+++C++ I+++EG+R+L+RGLG+N+++++P +AI+F+ Y
Query        53 GASVNRVVSPGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAY   102

HMM             EFMKeMFiDyfgeddnyWmWFwmnYMaGs*
                 +KE ++D F++ D++++++ + +MAG+
Query       103 SNCKEKLNDVFDP-DSTQVHMISAAMAGM     130
```

DKFZphtes3_21d4
----

group: signal transduction

DKFZphtes3_21d4 encodes a novel 464 amino acid putative GTP exchanging factor related to RCC1.

RCC1 (regulator of chromosome condensation) is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein. RCC1 promotes the exchange of bound GDP with GTP, acting as a guanine-nucleotide dissociation stimulator.

The new protein can find application in the regulation of gene expression by activition of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat.

similarity to RCC1-like G exchanging factor RLG

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="20"

Insert length: 2321 bp
Poly A stretch at pos. 2293, polyadenylation signal at pos. 2262

```
   1 GGGTCACGCA AGATGGCGGC GCCCAGAGGC TGCTGAGGCG CGGAACGGAG
  51 GATGGCGCTG GTGGCGTTGG TGGCTGGGGC TCGGCTGGGG CGGCGGCTGA
 101 GCGGGCCGGG GCTGGGGCGA GGGCACTGGA CGGCGGCCAG GCGCTCCCGG
 151 AGCCGGCGCG AAGCGGCAGA AGCCGAGGCG GAGGTGCCCG TGGTCCAGTA
 201 CGTGGGCGAG CGCGCTGCCC GCGCCGATCG CGTCTTCGTG TGGGGCTTCA
 251 GCTTCTCGGG GGCGCTGGGC GTGCCTTCCT TTGTGGTGCC CAGCTCCGGG
 301 CCCGGGCCCC GCGCCGGCGC CCGACCGCGC CGCAGGATCC AGCCCGTGCC
 351 CTATCGCCTG GAGCTGGACC AAAAGATTTC ATCTGCTGCT TGCGGCTATG
 401 GATTCACACT GCTGTCCTCT AAGACTGCGG ATGTTACGAA AGTCTGGGGG
 451 ATGGGACTCA ACAAAGATTC TCAGCTTGGA TTTCACAGGA GCCGGAAAGA
 501 TAAAACGAGG GGCTACGAGT ATGTGTTGGA GCCCTCACCC GTCTCCCTGC
 551 CTCTGGACAG ACCTCAGGAG ACACGGGTGC TGCAGGTCTC CTGCGGCCGA
 601 GCTCACTCTC TTGTGTTGAC TGACAGGGAA GGAGTCTTCA GCATGGGAAA
 651 CAATTCTTAT GGGCAATGTG GAAGAAAGGT GGTCGAAAAT GAAATTTACA
 701 GTGAAAGTCA CAGAGTCCAC AGGATGCAGG ACTTCGATGG CCAGGTGGTC
 751 CAGGTCGCCT GTGGTCAGGA TCATAGTCTG TTCCTGACGG ATAAAGGAGA
 801 AGTCTATTCT TGTGGATGGG GTGCTGATGG GCAAACAGGT CTGGGTCACT
 851 ACAATATCAC CAGCTCGCCC ACCAAGCTGG GTGGAGACCT GGCGGGAGTG
 901 AACGTTATCC AAGTTGCCAC CTACGGTGAT TGCTGCCTGG CCGTGTCCGC
 951 CGACGGAGGA CTTTTTGGTT GGGGAAACTC GGAGTACCTG CAGCTGGCCT
1001 CTGTCACTGA CTCCACACAG GTGAATGTGC CCCGCTGCTT ACACTTCTCA
1051 GGAGTGGGGA AGGTGCGACA GGCTGCATGC GGTGGCACGG GCTGTGCAGT
1101 GTTAAACGGA GAAGGACATG TTTTTGTCTG GGGCTATGGA ATTCTTGGGA
1151 AAGGTCCAAA CCTAGTGGAA AGTGCCGTCC CTGAAATGAT TCCACCCACT
1201 CTCTTTGGCT TGACGGAGTT CAACCCAGAA ATCCAGGTTT CCCGCATCCG
1251 ATGTGGACTC AGCCACTTTG CTGCACTGAC CAACAAAGGA GAGCTGTTTG
1301 TATGGGGCAA GAACATCCGA GGGTGCCTGG GAATCGGTCG CCTGGAGGAC
1351 CAGTATTTCC CATGGAGGGT GACGATGCCT GGGGAGCCTG TGGACGTGGC
1401 ATGTGGCGTG GACCACATGG TGACCCTGGC CAAGTCATTC ATCTAAACCT
1451 CCCTCACCTG CTTGGGCGGC CCCGTCCCGG GAACCACTGG CACTCCTTGG
1501 CAGAGGCCAG CGCGTGGCCA GCCCCCCGGG GTTCTTGGAT GGTGGTGGCG
1551 GAGGACCCTG CGTGCAGTGT GACGCTCTGT CCTGAATCCC TTAGCGGGTA
1601 CCTACCAGGA GGATCAGGGC AAGGTCCCTC TCCAGCTGCA GGTGAGGCCT
1651 GCGGAACTCA GCTTGGATGG CAGCCTTTGG TGGGCCGCTG TGGCCCGCAC
1701 GTCTCTGTTC TCTCCAAGTA ACATGCGACG GTGTCTGGTG TCACGTCTCG
1751 CCTGAGAAGC CCGTCTTAGG AAAGCTTAGC TTGAACACAG TGCTCGGGAG
1801 GTTTCTGCTC TGTCTGTCAT GGCAGTCTCT TGGTTTGTGT CTGGCCAAGG
1851 CCATGCGTGT GCCTCGGACC GAGCCCCAGC TTAGGCGAGG GAGTCAGGCT
1901 GGCTTCGGCC CTCGGTTTTC ATTCAGGCCA CCCTGCTCAT GGCCCTTCCT
1951 GGCCGCCTGC CACACCGCAA GCTCGCTGGG GGGACACTAG AAGCACCGTG
2001 GCCTGGGATT CCATCTGGAG CTGTCCGCAG GCACCAGCCC CAGCCTCCCA
2051 CCACGCTCAC TGCCTGGCTT GGAAAAGTTA AGAAGCCCCT CAGGAAGAGA
2101 ATCGAGGCTA AGTTCCTCTG CGCCGAGGGC CCCGAGCATA TCCGCCAAGG
2151 CTCAGCTGCA GTGCCAGGCG GAGGAGGAAG ATCCAGAAAT TGTGAACAAT
2201 GTTTGATTTA GTAGCGTGAC TTGCCTTTCC CTTTAAAAAC ATCTTTTACA
2251 AATCTGTCTT GGAATAAAGT CTATTTTCTG CCTTTTGGTT TTTAAAAAAA
2301 AAAAAAAAAA AAAAAAAAAA A
```

BLAST Results

Entry HS203358 from database EMBL:
human STS SHGC-31781.
Score = 1748, P = 1.1e-72, identities = 376/394

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 52 bp to 1443 bp; peptide length: 464
Category: similarity to known protein

```
  1 MALVALVAGA RLGRRLSGPG LGRGHWTAAR RSRSRREAAE AEAEVPVVQY
 51 VGERAARADR VFVWGFSFSG ALGVPSFVVP SSGPGPRAGA RPRRRIQPVP
101 YRLELDQKIS SAACGYGFTL LSSKTADVTK VWGMGLNKDS QLGFHRSRKD
151 KTRGYEYVLE PSPVSLPLDR PQETRVLQVS CGRAHSLVLT DREGVFSMGN
201 NSYGQCGRKV VENEIYSESH RVHRMQDFDG QVVQVACGQD HSLFLTDKGE
251 VYSCGWGADG QTGLGHYNIT SSPTKLGGDL AGVNVIQVAT YGDCCLAVSA
301 DGGLFGWGNS EYLQLASVTD STQVNVPRCL HFSGVGKVRQ AACGGTGCAV
351 LNGEGHVFVW GYGILGKGPN LVESAVPEMI PPTLFGLTEF NPEIQVSRIR
401 CGLSHFAALT NKGELFVWGK NIRGCLGIGR LEDQYFPWRV TMPGEPVDVA
451 CGVDHMVTLA KSFI
```

BLASTP hits

Entry CEW09G3_5 from database TREMBLNEW:
gene: "W09G3.3"; Caenorhabditis elegans cosmid W09G3
Score = 395, P = 9.3e-37, identities = 111/330, positives = 165/330

Entry Y032_HUMAN from database SWISSPROT:
HYPOTHETICAL PROTEIN KIAA0032.
Score = 309, P = 1.0e-24, identities = 96/308, positives = 143/308

Entry B38919 from database PIR:
hypothetical protein 2 - human (fragment)
Score = 309, P = 1.0e-24, identities = 96/308, positives = 143/308

Entry AF060219_1 from database TREMBLNEW:
product: "RCC1-like G exchanging factor RLG"; Homo sapiens RCC1-like G exchanging factor RLG mRNA, complete cds.
Score = 273, P = 4.0e-21, identities = 84/262, positives = 124/262

Entry S71752 from database PIR:
giant protein p619 - human
Score = 282, P = 1.1e-19, identities = 86/287, positives = 144/287

Alert BLASTP hits for DKFZphtes3_21d4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_21d4, frame 1

Report for DKFZphtes3_21d4.1

[LENGTH] 464
[MW] 49997.08
[pI] 8.74
[HOMOL] TREMBL:CEW09G3_5 gene: "W09G3.3"; Caenorhabditis elegans cosmid W09G3 5e-34

[FUNCAT] 04.07 rna transport [S. cerevisiae, YGL097w] 2e-09
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YGL097w] 2e-09
[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YGL097w] 2e-09
[FUNCAT] 04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S. cerevisiae, YGL097w] 2e-09
[FUNCAT] 04.01.04 rrna processing [S. cerevisiae, YGL097w] 2e-09
[FUNCAT] 04.03.03 trna processing [S. cerevisiae, YGL097w] 2e-09
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YGL097w] 2e-09

[FUNCAT]	30.04 organization of cytoskeleton [S. cerevisiae, YAL020c] 4e-06
[BLOCKS]	BL00870I
[BLOCKS]	BL00625B Regulator of chromosome condensation (RCC1) proteins
[BLOCKS]	BL00625A Regulator of chromosome condensation (RCC1) proteins
[PIRKW]	blocked amino end 3e-16
[PIRKW]	nucleus 3e-16
[PIRKW]	duplication 4e-08
[PIRKW]	tandem repeat 3e-16
[PIRKW]	DNA binding 3e-16
[PIRKW]	mitosis 3e-16
[PIRKW]	leucine zipper 3e-21
[SUPFAM]	pheromone response pathway component SRM1 4e-08
[SUPFAM]	WD repeat homology 3e-21
[PROSITE]	MYRISTYL 7
[PROSITE]	RCC1_2 2
[PROSITE]	AMIDATION 2
[PROSITE]	CAMP_PHOSPHO_SITE 1
[PROSITE]	CK2_PHOSPHO_SITE 5
[PROSITE]	TYR_PHOSPHO_SITE 2
[PROSITE]	GLYCOSAMINOGLYCAN 3
[PROSITE]	PKC_PHOSPHO_SITE 7
[PROSITE]	ASN_GLYCOSYLATION 2
[PFAM]	Regulator of chromosome condensation (RCC1)
[KW]	All_Beta
[KW]	LOW_COMPLEXITY 13.58 %

```
SEQ  MALVALVAGARLGRRLSGPGLGRGHWTAARRSRSRREAAEAEAEVPVVQYVGERAARADR
SEG  .xxxxxxxxxxxxxxxxxxxxxx...xxxxxxxxxxxxxxxxx.................
PRD  ccchhhhhhhhheeecccccccccchhhhhhhhhhhhhhhhhhceeeeeehhhhhhhhhh

SEQ  VFVWGFSFSGALGVPSFVVPSSGPGPRAGARPRRRIQPVPYRLELDQKISSAACGYGFTL
SEG  ............xxxxxxxxxxxxxxxxxxxxxx..........................
PRD  eeeecccccccccccceeeeecccccccccccccccccccccchhhhhhheeecccceee

SEQ  LSSKTADVTKVWGMGLNKDSQLGFHRSRKDKTRGYEYVLEPSPVSLPLDRPQETRVLQVS
SEG  ............................................................
PRD  eecccccceeeeccccccccccccccccccccccccceeeeeccccccccccccceeeee

SEQ  CGRAHSLVLTDREGVFSMGNNSYGQCGRKVVENEIYSESHRVHRMQDFDGQVVQVACGQD
SEG  ............................................................
PRD  cccceeeeeeccceeeeeeccccccccccccccccccccccccccccceeeeeeeecccc

SEQ  HSLFLTDKGEVYSCGWGADGQTGLGHYNITSSPTKLGGDLAGVNVIQVATYGDCCLAVSA
SEG  ............................................................
PRD  eeeeeecccceeeecccccccccccccccccccccccccccceeeeeeeecccceeeeeec

SEQ  DGGLFGWGNSEYLQLASVTDSTQVNVPRCLHFSGVGKVRQAACGGTGCAVLNGEGHVFVW
SEG  ............................................................
PRD  ccceeeeccccccccccccccccccccccccccccccceeeeecccceeeeeecccceeee

SEQ  GYGILGKGPNLVESAVPEMIPPTLFGLTEFNPEIQVSRIRCGLSHFAALTNKGELFVWGK
SEG  ............................................................
PRD  cccccccccccccccccccccceeeeeeecccceeeeeeeecccceeeeeecccceeeecc

SEQ  NIRGCLGIGRLEDQYFPWRVTMPGEPVDVACGVDHMVTLAKSFI
SEG  ............................................
PRD  cccccccccccccccccceeecccceeeeecccccccccccccc
```

Prosite for DKFZphtes3_21d4.1

PS00001	200->204	ASN_GLYCOSYLATION	PDOC00001
PS00001	268->272	ASN_GLYCOSYLATION	PDOC00001
PS00002	17->21	GLYCOSAMINOGLYCAN	PDOC00002
PS00002	82->86	GLYCOSAMINOGLYCAN	PDOC00002
PS00002	333->337	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	14->18	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	34->37	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	190->193	PKC_PHOSPHO_SITE	PDOC00005
PS00005	219->222	PKC_PHOSPHO_SITE	PDOC00005
PS00005	246->249	PKC_PHOSPHO_SITE	PDOC00005
PS00005	410->413	PKC_PHOSPHO_SITE	PDOC00005
PS00006	34->38	CK2_PHOSPHO_SITE	PDOC00006
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	290->294	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006

PS00007	209->217	TYR_PHOSPHO_SITE	PDOC00007
PS00007	208->217	TYR_PHOSPHO_SITE	PDOC00007
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	238->244	MYRISTYL	PDOC00008
PS00008	277->283	MYRISTYL	PDOC00008
PS00008	302->308	MYRISTYL	PDOC00008
PS00008	344->350	MYRISTYL	PDOC00008
PS00009	12->16	AMIDATION	PDOC00009
PS00009	206->210	AMIDATION	PDOC00009
PS00626	179->190	RCC1_2	PDOC00544
PS00626	235->246	RCC1_2	PDOC00544

Pfam for DKFZphtes3_21d4.1

```
HMM_NAME     Regulator of chromosome condensation (RCC1)

HMM                *IAaGqHHTVCLTqDGRVYtWG*
                    +A GQ+H++ LT++G VY++G
Query        235    VACGQDHSLFLTDKGEVYSCG    255
```

DKFZphtes3_21j15
----

group: transcription factors

DKFZphtes3_21j15 encodes a novel 898 amino acid protein with similarity human NY-CO-33 protein.

NY-CO-33 is a protein recognised by autologous antibodies of human colon cancer patients. The novel protein contains 4 C2H2 Zinc fingers and is a new putativ transcription factor.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

strong similarity to "NY-CO-33"

complete cDNA, complete cds, potential start at bp 27, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 4407 bp
Poly A stretch at pos. 4321, polyadenylation signal at pos. 4301

```
   1 CGCTGCAGCA GGTGTCACAG AGCCGCATGC TCCCGGAGCC CAGCCTCTTC
  51 AGCACCGTGC AGCTGTACCG GCAGAGCAGC AAGCTCTATG GCTCCATCTT
 101 CACGGGGGCC AGCAAGTTCC GCTGTAAGGA CTGCAGCGCT GCCTACGACA
 151 CCCTGGTGGA GTTGACAGTG CACATGAACG AGACGGGGCA TTACCGCGAC
 201 GACAACCATG AGACCGATAA CAACAACCCC AAGCGCTGGT CCAAGCCTCG
 251 CAAACGCTCC TTGCTGGAAA TGGAAGGGAA GGAAGACGCC CAGAAGGTGC
 301 TGAAGTGCAT GTACTGTGGC CACTCCTTTG AGTCCCTGCA GGATTTGAGT
 351 GTCCATATGA TCAAAACAAA ACACTACCAA AAAGTGCCTC TGAAGGAACC
 401 CGTCACTCCT GTCGCCGCCA AAATCATCCC TGCCACTCGG AAGAAAGCTT
 451 CCCTGGAGCT GGAGCTCCCC AGCTCCCCAG ATTCCACAGG TGGAACCCCC
 501 AAAGCCACCA TCTCAGACAC CAACGATGCA CTTCAGAAGA ACTCCAACCC
 551 TTACATCACG CCAAATAATC GGTACGGCCA CCAGAATGGG GCCAGCTATG
 601 CATGGCACTT TGAGGCCCGG AAGTCGCAGA TCCTGAAGTG CATGGAGTGT
 651 GGGAGCTCGC ATGACACCCT GCAGGAGCTC ACTGCCCACA TGATGGTCAC
 701 TGGCCACTTC ATCAAGGTCA CCAACTCTGC TATGAAAAAG GGGAAGCCCA
 751 TTGTGGAGAC GCCTGTCACA CCTACCATCA CAACCCTGCT GGATGAGAAG
 801 GTCCAGTCCG TGCCCCTGGC AGCCACCACC TTCACGTCCC CCTCCAATAC
 851 ACCTGCCAGC ATCTCCCCAA AACTGAATGT GGAGGTCAAG AAGGAAGTCG
 901 ACAAGGAGAA AGCGGTCACT GACGAGAAAC CTAAGCAAAA AGACAAGCCT
 951 GGCGAAGAAG AGGAGAAGTG TGACATCTCT TCCAAATACC ATTACTTGAC
1001 TGAAAATGAC TTAGAAGAGA GTCCCAAGGG GGGGCTTGAT ATCCTCAAAT
1051 CCTTGGAAAA CACAGTGACA TCCGCAATCA ACAAGGCCCA GAACGGCACT
1101 CCTAGCTGGG GGGGCTATCC CAGCATCCAT GCCGCCTACC AACTTCCCAA
1151 CATGATGAAG TTGTCCCTGG GCTCGTCGGG GAAGAGCACG CCCCTGAAAC
1201 CCATGTTTGG CAACAGTGAG ATTGTCTCCC CGACGAAAAA CCAGACCCTG
1251 GTCTCTCCAC CCAGCAGCCA GACGTCCCCC ATGCCCAAGA CAAACTTTCA
1301 TGCCATGGAG GAGCTGGTGA AAAAGGTCAC TGAGAAAGTT GCCAAAGTGG
1351 AGGAGAAGAT GAAGGAGCCG GATGGGAAGC TTTCCCCGCC CAAGCGGGCC
1401 ACTCCCTCCC CATGTAGCAG CGAAGTCGGG GAACCCATCA AGATGGAGGC
1451 ATCCAGCGAT GGGGGCTTCC GCAGCCAGGA GAACAGCCCC AGCCCCCCGC
1501 GGGATGGGTG CAAGGATGGG AGCCCCCTCG CTGAGCCGGT GGAGAATGGC
1551 AAGGAGCTGG TGAAGCCCCT AGCCAGCAGT TTGAGTGGCA GCACGGCCAT
1601 CATCACCGAC CACCCGCCTG AACAGCCTTT TGTTAACCCT TTGAGCGCCC
1651 TGCAGTCAGT CATGAACATT CACCTGGGCA AGGCCGCCAA GCCCTCCCTG
1701 CCTGCCCTGG ACCCCATGAG CATGCTTTTC AAGATGAGCA ACAGCCTGGC
1751 GGAGAAGGCT GCTGTGGCCA CCCCGCCGCC CCTGCAGTCC AAGAAGGCAG
1801 ACCACCTCGA CCGCTATTTC TACCACGTCA ACAACGACCA GCCCATAGAC
1851 TTGACAAAAG GGAAGAGTGA CAAAGGCTGC TCCTTGGGTT CAGTGCTTCT
1901 GTCACCCACG TCCACAGCCC CGGCAACCTC CTCATCCACG GTGACAACGG
1951 CAAAGACATC TGCCGTCGTA TCATTCATGT CAAACTCGCC GCTACGCGAG
2001 AATGCCTTGT CAGATATATC CGATATGCTG AAGAACTTGA CAGAGAGCCA
2051 CACGTCAAAA TCCTCCACTC CTTCCAGCAT CTCCGAGAAG TCTGACATTG
2101 ACGGGGCCAC TCTGGAGGAG GCTGAGGAGT CGACGCCCGC CCAGAAGAGG
2151 AAGGGCCGCC AGTCAAACTG GAACCCCCAG CACCTCCTGA TCCTCCAGGC
2201 CCAGTTTGCC GCCAGCCTCC GGCAGACCTC AGAAGGGAAG TACATCATGT
2251 CAGACCTGAG CCCCCAGGAG CGGATGCATA TCTCCAGGTT CACCGGGCTG
2301 TCCATGACCA CCATCAGCCA CTGGCTGGCC AACGTGAAAT ACCAGCTTCG
2351 AAGGACAGGT GGAACAAAGT TCCTCAAAAA CTTGGACACT GGCCACCCCG
2401 TCTTCTTTTG TAACGATTGT GCGTCCCAAA TCAGGACTCC TTCCACGTAC
2451 ATCAGTCACC TAGAGTCACA CTTAGGCTTC CGGCTACGGG ACTTATCCAA
2501 ACTGTCCACC GAACAGATTA ACAGTCAGAT AGCACAAACC AAGTCACCGT
2551 CAGAAAAAAT GGTGACGTCC TCCCCCGAGG AAGACCTGGG GACTTCCTAT
2601 CAGTGCAAAC TTTGCAATCG GACCTTTGCC AGCAAGCACG CTGTTAAACT
```

```
2651 TCACCTTAGC AAAACACACG GGAAATCTCC GGAAGACCAC CTTCTGTATG
2701 TCTCTGAGTT AGAGAAGCAG TAGCATTTGC TTTTGATAGA AAGGACTGCA
2751 GTTTGCTTTG AGGGAAACTG TGGAAGGCAC CTTCAGGCCC CCTCTGACTT
2801 GTTGTTCTTG GCACATGTTC TTATTTTAAC TGCAGAGAAT CACTCTGGGC
2851 TGGACTGTTT TGTATAACTG TACAGTGTTT AATAGAGGTG CATAATCAGC
2901 TGTTGTTACT GGTAAAAATAT GAAGGTTAAA ATGCAGTGGT AAGTGTTTGG
2951 AACTTTGTGT AAACGGGATT TAGTTGTGAG CATCCTCCCG ATGCTTCAAG
3001 CTGCATGCAT TAACAGACAG TTTAATTAAG CATTTATAAC GGAATCAGGC
3051 ACACCTTTTC CACGAGACTC GAGTGTGCTG GCATTTCTCA CCCTTTCATC
3101 TTTAGCCCTC TGAGTACTTT GAAGCACTTT TGCATTAATT TGGTTAAAAA
3151 ATAAAATAAA ATAATAATAA TGTATGAAGC TCTGTTTTTT AAACTCCTTA
3201 CCAGCTTAGT TATAATGAAT AATATGAACC TCCATTTATG CAGGTCTGCA
3251 GGGGTATAAC ACGCCTTGAA ATTTAAAAGA ATATTATTTT CACATTGAAA
3301 CATAGATGTA TATATTGTAT AGATTTCAGA CTCTCTTATG AAAAAAAATG
3351 TGATTGTGGT TAAATGACCT TTTTCTTGCA TTTATAGCAA CAGTGTTTTA
3401 TGCACCTGCT ATGCTCTGGG CATAAGCTGT GCCTATGTAT AGTGTATATT
3451 TCTTTTTTTC TTTTTTTTAA GGTCTATGGG TTTTGTTTTT TACATGCAAA
3501 CATTGTAAAT TATACAGAAG ATACCACAGA TAGCATTTAT AAAGTATACA
3551 GAAACATTAT CTGAAAGCAA AGTATGATAG TTTGTTTTGC TATACAGTAC
3601 ATCTATATTG ATAGAGGTTC ATGTTTAAAT TATACATATT TATTAGCATC
3651 ATATTGTCAT TTGTTTTGAG CAGTCTGAAT AAACGAGACC GGGAAAGACA
3701 TCCCTGGCAG GCATCAGAAC TATTTTGCAC ATGATTTTTA AAGGTATTTA
3751 TTAGAAATCA AAGAACACTC AAAATAAACT CAGTGCTCAA AGGGTTAAGT
3801 CTATTTGAAA AGGTTAAAAA AAAGAACAAA AAAAAAAAAA GAACTTGTAC
3851 TGTATTTCCT AAACATTGAT AAAGCCTTTA AAATGTTTGT ACTGTAATAC
3901 TTTGCTTAAA AGTCATGAGG CATTCTGTGA TCCAACCTCT TTCACTTATT
3951 TATAAGCCCT CTTGGTTGCT ATTCCATATT GTAGGATGCC TTTCTATTTC
4001 AATTGGTAAC TTTCTGTTTT GTTCTTCCTA ATTATTCTCC CAAGATCCCA
4051 CACTGCAGCT TTATCTTTAG GCTTATGAAA GGTAACCCGT GGTTACCGGC
4101 TCTCCAAGTG ATTCTGTTCT TCTCCATTTT TGGCAGTTAA TTTGCAGAAG
4151 TAACTGACAG CTGACACCAT ATGAGAACCT TTGTATAAAA TATTGGCATG
4201 TAAACAGCAC AGACACCGTA ACACACTCTG TGCCCTGTTT GGTTGTTGAC
4251 AATGAAGCAC CATTATGTGA CTCTTCATAT AACCCTTTTT TCTACGGCAG
4301 CATTAAAATT GTCTTTTTGC TATAAAAAAA AAAAAAAAAA AAAAAAAAAA
4351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
4401 AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 27 bp to 2720 bp; peptide length: 898
Category: strong similarity to known protein

```
  1 MLPEPSLFST VQLYRQSSKL YGSIFTGASK FRCKDCSAAY DTLVELTVHM
 51 NETGHYRDDN HETDNNNPKR WSKPRKRSLL EMEGKEDAQK VLKCMYCGHS
101 FESLQDLSVH MIKTKHYQKV PLKEPVTPVA AKIIPATRKK ASLELELPSS
151 PDSTGGTPKA TISDTNDALQ KNSNPYITPN NRYGHQNGAS YAWHFEARKS
201 QILKCMECGS SHDTLQELTA HMMVTGHFIK VTNSAMKKGK PIVETPVTPT
251 ITTLLDEKVQ SVPLAATTFT SPSNTPASIS PKLNVEVKKE VDKEKAVTDE
301 KPKQKDKPGE EEEKCDISSK YHYLTENDLE ESPKGGLDIL KSLENTVTSA
351 INKAQNGTPS WGGYPSIHAA YQLPNMMKLS LGSSGKSTPL KPMFGNSEIV
401 SPTKNQTLVS PPSSQTSPMP KTNFHAMEEL VKKVTEKVAK VEEKMKEPDG
451 KLSPPKRATP SPCSSEVGEP IKMEASSDGG FRSQENSPSP PRDGCKDGSP
501 LAEPVENGKE LVKPLASSLS GSTAIITDHP PEQPFVNPLS ALQSVMNIHL
551 GKAAKPSLPA LDPMSMLFKM SNSLAEKAAV ATPPPLQSKK ADHLDRYFYH
601 VNNDQPIDLT KGKSDKGCSL GSVLLSPTST APATSSSTVT TAKTSAVVSF
651 MSNSPLRENA LSDISDMLKN LTESHTSKSS TPSSISEKSD IDGATLEEAE
701 ESTPAQKRKG RQSNWNPQHL LILQAQFAAS LRQTSEGKYI MSDLSPQERM
751 HISRFTGLSM TTISHWLANV KYQLRRTGGT KFLKNLDTGH PVFFCNDCAS
801 QIRTPSTYIS HLESHLGFRL RDLSKLSTEQ INSQIAQTKS PSEKMVTSSP
851 EEDLGTSYQC KLCNRTFASK HAVKLHLSKT HGKSPEDHLL YVSELEKQ
```

BLASTP hits

No BLASTP hits available

```
           Alert BLASTP hits for DKFZphtes3_21j15, frame 3

TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33";  Homo
sapiens antigen NY-CO-33 (NY-CO-33) mRNA, complete cds., N = 1, Score =
1039, P = 5.5e-105

PIR:A38437 probable homeotic protein tsh - fruit fly (Drosophila
melanogaster), N = 3, Score = 158, P = 7.2e-09

TREMBL:CE33058_1 gene: "unc-89"; product: "UNC-89";  Caenorhabditis
elegans UNC-89 (unc-89) gene, complete cds., N = 2, Score = 175, P =
3.3e-07


>TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33";  Homo
     sapiens antigen NY-CO-33 (NY-CO-33) mRNA, complete cds.
          Length = 687

  HSPs:

 Score = 1039 (155.9 bits), Expect = 5.5e-105, P = 5.5e-105
 Identities = 244/504 (48%), Positives = 319/504 (63%)

Query:   170 QKNSNPYITPNNRYGHQNGASYAWHFEARKSQILKCMECGSSHDTLQELTAHMMVTGHFI 229
             QK +NPY+TPNNRYG+QNGASY W FEARK+QILKCMECGSSHDTLQ+LTAHMMVTGHF+
Sbjct:    14 QKAANPYVTPNNRYGYQNGASYTWQFEARKAQILKCMECGSSHDTLQQLTAHMMVTGHFL 73

Query:   230 KVTNSAMKKGKPIVETPVTPTITTLLDEKVQSVPLAATTFTS-PSNT----PASISPKLN 284
             KVT SA KKGK +V  PV        ++EK+QS+PL  TT T  P+++     P S +
Sbjct:    74 KVTTSASKKGKQLVLDPV-------VEEKIQSIPLPPTTHTRLPASSIKKQPDSPAGSTT 126

Query:   285 VEVKKEVDKEKA-VTDEKPKQKDKPGEEEEKCDISSKYHYLTENDLEESPKGGLDILKSL 343
              E KKE +KEK  V  +  K K++  +  EK + S+ Y YL E DL++SPKGGLDILKSL
Sbjct:   127 SEEKKEPEKEKPPVAGDAEKIKEESEDSLEKFEPSTLYPYLREEDLDDSPKGGLDILKSL 186

Query:   344 ENTVTSAINKAQNGTPSWGGYPSIHAAYQLPNMMKLSLGSSGKSTPLKPMF-GNSEIVSP 402
             ENTV++AI+KAQNG PSWGGYPSIHAAYQLP  +K  L ++ +S  ++P + G  + +S
Sbjct:   187 ENTVSTAISKAQNGAPSWGGYPSIHAAYQLPGTVK-PLPAAVQSVQVQPSYAGGVKSLSS 245

Query:   403 TKNQTLVSPPSSQTSPMPKTNFHAMEELVKKVTEKV-AKVEEKMKEPDGKLSPPKRATPS 461
              ++  L+  P S T P  K+N  AMEELV+KVT KV  K EE+  E + K S  K A  S
Sbjct:   246 AEHNALLHSPGSLTPPPHKSNVSAMEELVEKVTGKVNIKKEERPPEKE-KSSLAKAA--S 302

Query:   462 PCSSEVGEPIKMEASSDGGFRSQENSPSPPRDGCKDGSPLAEPVENGKELVKPLASSLSG 521
             P + E  + K E  S    + Q+  P            K  PL      NG E +K   ++
Sbjct:   303 PIAKENKDFPKTEEVSG---KPQKKGPEAETWEAKKEGPLDVHTPNGTEPLKAKVTNGCN 359

Query:   522 STAIITDHPPEQPFVNPLSALQSVMNIHLGKAAKPSLPALDPMSMLFKMSNSLAEKAAVA 581
             +  II DH PE  F+NPLSALQS+MN HLGK +KP  P+LDP++ML+K+SNS+ +K
Sbjct:   360 NLGIIMDHSPEPSFINPLSALQSIMNTHLGKVSKPVSPSLDPLAMLYKISNSMLDKPVYP 419

Query:   582 TPPPLQSKKADHLDRYFYHVNNDQPIDLTKGKSDK-GCSLGSVLLSPTSTAPATSSSTVT 640
              P     K+AD +DRY+Y  N+DQPIDLTK K+      S+   + SP    +      S +
Sbjct:   420 ATPV---KQADAIDRYYYE-NSDQPIDLTKSKNKPLVSSVADSVASPLRESALMDISDMV 475

Query:   641 TAKTSAVVSFMSN-SPLRENALSDISDMLKNLTE 673
               T  +    S  S + E + +D S   + L E
Sbjct:   476 KNLTGRLTPKSSTPSTVSEKSDADGSSFEEALDE 509

 Score = 865 (129.8 bits), Expect = 7.4e-95, P = 7.4e-95
 Identities = 211/434 (48%), Positives = 268/434 (61%)

Query:   447 EPDGKLSPPKRATPSPCSSEVG--EPIKMEASSDGGFRSQENSPSPPRDG-CKDGSPLAE 503
             E + L  P  TP P S V   E +  + +         + +E  P    +  K  SP+A+
Sbjct:   247 EHNALLHSPGSLTPPPHKSNVSAMEELVEKVTGKVNIKKEERPPEKEKSSLAKAASPIAK 306

Query:   504 -----P-VE--NGKELVK-PLASSLSGSTAIITD-HPPE--QPFVNPLSALQSVMNIHLG 551
                  P  E +GK   K  P A +            D H P  +P   ++  +  + I +
Sbjct:   307 ENKDFPKTEEVSGKPQKKGPEAETWEAKKEGPLDVHTPNGTEPLKAKVTNGCNNLGIIMD 366

Query:   552 KAAKPSLPALDPMSMLFKMSNSLAEKAAVATPPPLQSKKADHLDRYFYHVNN---DQPID 608
             + +PS     ++P+S L  + N+   K +    P L    D L     Y ++N  D+P+
Sbjct:   367 HSPEPSF--INPLSALQSIMNTHLGKVSKPVSPSL-----DPL-AMLYKISNSMLDKPV- 417

Query:   609 LTKGKSDKGCSLGSVLLSPTSTAPATSSSTVTTAKTSAVVSFMSNSPLRENALSDISDML 668
                       K             S  P  + +    S+V  ++ SPLRE+AL DISDM+
Sbjct:   418 -YPATPVKQADAIDRYYYENSDQPIDLTKSKNKPLVSSVADSVA-SPLRESALMDISDMV 475

Query:   669 KNLTESHTSKSSTPSSISEKSDIDGATLEEA-EESTPAQKRKGRQSNWNPQHLLILQAQF 727
             KNLT    T KSSTPS++SEKSD DG++ EEA +E +P   KRKGRQSNWNPQHLLILQAQF
```

```
Sbjct:   476 KNLTGRLTPKSSTPSTVSEKSDADGSSFEEALDELSPVHKRKGRQSNWNPQHLLILQAQF 535

Query:   728 AASLRQTSEGKYIMSDLSPQERMHISRFTGLSMTTISHWLANVKYQLRRTGGTKFLKNLD 787
             A+SLR+T+EGKYIMSDL PQER+HIS+FTGLSMTTISHWLANVKYQLRRTGGTKFLKNLD
Sbjct:   536 ASSLRETTEGKYIMSDLGPQERVHISKFTGLSMTTISHWLANVKYQLRRTGGTKFLKNLD 595

Query:   788 TGHPVFFCNDCASQIRTPSTYISHLESHLGFRLRDLSKLSTEQINSQIAQTKSPSEKMV- 846
             TGHPVFFCNDCASQ RT STYISHLE+HLGF L+DLSKL   QI  Q   +K  + K +
Sbjct:   596 TGHPVFFCNDCASQFRTASTYISHLETHLGFSLKDLSKLPLNQIQEQQNVSKVLTNKTLG 655

Query:   847 -TSSPEEDLGTSYQCKLCNRTFASK 870
               + EEDLG+++QCKLCNRTFA +
Sbjct:   656 PLGATEEDLGSTFQCKLCNRTFAKQ 680

 Score = 98 (14.7 bits), Expect = 7.4e-95, P = 7.4e-95
 Identities = 32/95 (33%), Positives = 47/95 (49%)

Query:    90 KVLKCMYCGHSFESLQDLSVHMIKTKHYQKVPL------KEPVT-PVAAKIIPATRKKAS 142
             ++LKCM CG S ++LQ L+ HM+ T H+ KV         K+ V  PV  + I +     +
Sbjct:    45 QILKCMECGSSHDTLQQLTAHMMVTGHFLKVTTSASKKGKQLVLDPVVEEKIQSIPLPPT 104

Query:   143 LELELPSS-----PDSTGGTPKATISDTNDALQKNSNP 175
                 LP+S     PDS G+   T S+     +K   P
Sbjct:   105 THTRLPASSIKKQPDSPAGS---TTSEEKKEPEKEKPP 139

 Score = 81 (12.2 bits), Expect = 4.6e-93, P = 4.6e-93
 Identities = 13/29 (44%), Positives = 20/29 (68%)

Query:    28 ASKFRCKDCSAAYDTLVELTVHMNETGHY 56
             A   +C +C +++DTL +LT HM  TGH+
Sbjct:    44 AQILKCMECGSSHDTLQQLTAHMMVTGHF 72
```

Pedant information for DKFZphtes3_21j15, frame 3

Report for DKFZphtes3_21j15.3

```
[LENGTH]      898
[MW]          98486.72
[pI]          8.61
[HOMOL]       TREMBL:AF039698_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens
antigen NY-CO-33 (NY-CO-33) mRNA, complete cds. 0.0
[BLOCKS]      BL00028 Zinc finger, C2H2 type, domain proteins
[PIRKW]       zinc finger 1e-06
[PIRKW]       DNA binding 1e-06
[PIRKW]       transcription regulation 1e-06
[PROSITE]     MYRISTYL           9
[PROSITE]     ZINC_FINGER_C2H2       4
[PROSITE]     CAMP_PHOSPHO_SITE      5
[PROSITE]     CK2_PHOSPHO_SITE       19
[PROSITE]     TYR_PHOSPHO_SITE       2
[PROSITE]     PKC_PHOSPHO_SITE       15
[PROSITE]     ASN_GLYCOSYLATION      4
[PFAM]        Zinc finger, C2H2 type
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY    11.36 %

SEQ    MLPEPSLFSTVQLYRQSSKLYGSIFTGASKFRCKDCSAAYDTLVELTVHMNETGHYRDDN
SEG    ............................................................
PRD    cccceeeeeeeeccccceeeeeeeeccccceeecccchhhhhhhhhhhcccccccccccc

SEQ    HETDNNNPKRWSKPRKRSLLEMEGKEDAQKVLKCMYCGHSFESLQDLSVHMIKTKHYQKV
SEG    ............................................................
PRD    cccccccccccccccchhhhhhhcchhhhhhhhhhcccccchhhhheeeeeeeecceeee

SEQ    PLKEPVTPVAAKIIPATRKKASLELELPSSPDSTGGTPKATISDTNDALQKNSNPYITPN
SEG    ....................xxxxxxxxxx..............................
PRD    eccccccceeeeeeehhhhhhhhhhcccccccccccccccceeeecccchhhhhcccccccc

SEQ    NRYGHQNGASYAWHFEARKSQILKCMECGSSHDTLQELTAHMMVTGHFIKVTNSAMKKGK
SEG    ............................................................
PRD    cccccccchhhhhhhhhhhhhhhhccccccccchhhhhhhhhhceeeeeecccccccccc

SEQ    PIVETPVTPTITTLLDEKVQSVPLAATTFTSPSNTPASISPKLNVEVKKEVDKEKAVTDE
SEG    xxxxxxxxxxxxx.................................xxxxxxxxxxxxxxx
PRD    cccccccccchhhhhhhhcccccccccccccccccccccccccccccccccchhhhhhcc

SEQ    KPKQKDKPGEEEEKCDISSKYHYLTENDLEESPKGGLDILKSLENTVTSAINKAQNGTPS
```

```
SEG     x...........................................................
PRD     ccccccccccccccchhhhhhhhhhhcccccccccchhhhhhhhhhhhhhhhhhcccccc

SEQ     WGGYPSIHAAYQLPNMMKLSLGSSGKSTPLKPMFGNSEIVSPTKNQTLVSPPSSQTSPMP
SEG     ............................................................
PRD     cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ     KTNFHAMEELVKKVTEKVAKVEEKMKEPDGKLSPPKRATPSPCSSEVGEPIKMEASSDGG
SEG     .......xxxxxxxxxxxxxxxxxxxx.................................
PRD     ccchhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccceeeeecccc

SEQ     FRSQENSPSPPRDGCKDGSPLAEPVENGKELVKPLASSLSGSTAIITDHPPEQPFVNPLS
SEG     ............................................................
PRD     cccccccccccccccccccccccccccccccccccccccccceeeeecccccccccccccc

SEQ     ALQSVMNIHLGKAAKPSLPALDPMSMLFKMSNSLAEKAAVATPPPLQSKKADHLDRYFYH
SEG     ............................................................
PRD     chhhhhhccccccccccccccchhhhhhhhhhhhhhcccccccccccccccccccceeee

SEQ     VNNDQPIDLTKGKSDKGCSLGSVLLSPTSTAPATSSSTVTTAKTSAVVSFMSNSPLRENA
SEG     .........................xxxxxxxxxxxxxxxxxxxxxxx............
PRD     eccccccceeecccccccccccceeeccccccccccccceeeeceeeeeeeecccccccchhh

SEQ     LSDISDMLKNLTESHTSKSSTPSSISEKSDIDGATLEEAEESTPAQKRKGRQSNWNPQHL
SEG     ...........xxxxxxxxxxxxxxxxxx...............................
PRD     hhhhhhhhhhhcccccccccccccceeeecccccchhhhhhhhccchhhhhhhcccccccchh

SEQ     LILQAQFAASLRQTSEGKYIMSDLSPQERMHISRFTGLSMTTISHWLANVKYQLRRTGGT
SEG     ............................................................
PRD     hhhhhhhhhhhhhccccceeeccccccchhhhhhhhcccccchhhhhhhhhhhhhhhhhcccc

SEQ     KFLKNLDTGHPVFFCNDCASQIRTPSTYISHLESHLGFRLRDLSKLSTEQINSQIAQTKS
SEG     ............................................................
PRD     ceeeccccccccceeeccccceeeeccccchhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhcc

SEQ     PSEKMVTSSPEEDLGTSYQCKLCNRTFASKHAVKLHLSKTHGKSPEDHLLYVSELEKQ
SEG     ..........................................................
PRD     cccceeeecccccccccccceeehhhhhhhhhhhhhhhhhhcccccccccccceeeeeeeecccc
```

Prosite for DKFZphtes3_21j15.3

PS00001	51->55	ASN_GLYCOSYLATION	PDOC00001
PS00001	405->409	ASN_GLYCOSYLATION	PDOC00001
PS00001	670->674	ASN_GLYCOSYLATION	PDOC00001
PS00001	864->868	ASN_GLYCOSYLATION	PDOC00001
PS00004	69->73	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	75->79	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	139->143	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	432->436	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	456->460	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	17->20	PKC_PHOSPHO_SITE	PDOC00005
PS00005	137->140	PKC_PHOSPHO_SITE	PDOC00005
PS00005	157->160	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	318->321	PKC_PHOSPHO_SITE	PDOC00005
PS00005	332->335	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	588->591	PKC_PHOSPHO_SITE	PDOC00005
PS00005	614->617	PKC_PHOSPHO_SITE	PDOC00005
PS00005	641->644	PKC_PHOSPHO_SITE	PDOC00005
PS00005	676->679	PKC_PHOSPHO_SITE	PDOC00005
PS00005	686->689	PKC_PHOSPHO_SITE	PDOC00005
PS00005	730->733	PKC_PHOSPHO_SITE	PDOC00005
PS00005	842->845	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	78->82	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	149->153	CK2_PHOSPHO_SITE	PDOC00006
PS00006	161->165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	253->257	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	684->688	CK2_PHOSPHO_SITE	PDOC00006
PS00006	689->693	CK2_PHOSPHO_SITE	PDOC00006
PS00006	695->699	CK2_PHOSPHO_SITE	PDOC00006
PS00006	745->749	CK2_PHOSPHO_SITE	PDOC00006

PS00006	810->814	CK2_PHOSPHO_SITE	PDOC00006
PS00006	840->844	CK2_PHOSPHO_SITE	PDOC00006
PS00006	848->852	CK2_PHOSPHO_SITE	PDOC00006
PS00006	884->888	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	732->740	TYR_PHOSPHO_SITE	PDOC00007
PS00007	883->892	TYR_PHOSPHO_SITE	PDOC00007
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	156->162	MYRISTYL	PDOC00008
PS00008	188->194	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00008	479->485	MYRISTYL	PDOC00008
PS00008	494->500	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	617->623	MYRISTYL	PDOC00008
PS00008	757->763	MYRISTYL	PDOC00008
PS00028	795->816	ZINC_FINGER_C2H2	PDOC00028
PS00028	860->882	ZINC_FINGER_C2H2	PDOC00028
PS00028	33->56	ZINC_FINGER_C2H2	PDOC00028
PS00028	94->117	ZINC_FINGER_C2H2	PDOC00028

Pfam for DKFZphtes3_21j15.3

```
HMM_NAME       Zinc finger, C2H2 type

HMM                  *CpwPDCgKtFrrwsNLrRHMR..T.H*
                      C++  C   ++ +   +L+ HM+    H
Query           33   CKD--CSAAYDTLVELTVHMNET-GH      55

26.69  (bits) f:   94 t:  116 Target: dkfzphtes3_21j15.3 strong similarity to "NY-CO-33"
  Alignment to HMM consensus:
Query                *CpwPDCgKtFrrwsNLrRHMR..T.H*
                      C +   CG +F +   +L  HM+    H
  dkfzphtes3    94   CMY--CGHSFESLQDLSVHMIKT-KH     116

Query          f:  795 t:  815 Target: dkfzphtes3_21j15.3 strong similarity to "NY-CO-33"
  Alignment to HMM consensus:
HMM                  *CpwPDCgKtFrrwsNLrRHMRTH*
                      C++  C      R++S+++ H+ +H
Query          795   CND--CASQIRTPSTYISHLESH    815

27.12  (bits) f:  860 t:  881 Target: dkfzphtes3_21j15.3 strong similarity to "NY-CO-33"
  Alignment to HMM consensus:
Query                *CpwPDCgKtFrrwsNLrRHMR.T.H*
                      C+    C++TF +++   + H+     H
  dkfzphtes3   860   CKL--CNRTFASKHAVKLHLSK-TH     881
```

DKFZphtes3_21116
----

group: intracellular transport and trafficking

DKFZphtes3_21116 encodes a novel 66 amino acid protein nearly identical to rat ribosome attached membrane protein 4 (ramp4).

The novel protein seems to be the human orthologe of rat ramp 4. Ramp4 is involved in the regulation of translocation of proteins into endoplasmic reticulum, e.g. of the MHC class II associated invariant (gamma) chain.

The new protein can find application in modulation of protein translocation into the endoplasmic reticulum.

identical to rat ribosome attached membrane protein 4

ORF Bp 316-513 (66 aa) see BLASTX

Sequenced by LMU

Locus: unknown

Insert length: 2488 bp
Poly A stretch at pos. 2464, polyadenylation signal at pos. 2442

```
   1 CTTCCTCTTT CACTCCGCGC TCACGGCGGC GGCCAAAGCG GCGGCGACGG
  51 CGGCGCGAGA ACGACCCGGC GGCCAGTTCT CTTCCTCCTG CGCACCTGCC
 101 CCGCTCGGTC AGTCAGTCGG CGGCCGGCGC CCGGCTTGTG CTCAGACCTC
 151 GCGCTTGCGG CGCCCAGGCC CAGCGGCCGT AGCTAGCGTC TGGCCTGAGA
 201 ACCTCGGCGC TCCGGCGGCG CGGGCACCAC GAGCCGAGCC TCGCAGCGGC
 251 TCCAGAGGAG GCAGGCGAGT GAGCGAGTCC GAGGGGTGGC CGGGGCAGGT
 301 GGTGGCGCCG CGAAGATGGT CGCCAAGCAA AGGATCCGTA TGGCCAACGA
 351 GAAGCACAGC AAGAACATCA CCCAGCGCGG CAACGTCGCC AAGACCTCGA
 401 GAAATGCCCC CGAAGAGAAG GCGTCTGTAG GACCCTGGTT ATTGGCTCTC
 451 TTCATTTTTG TTGTCTGTGG TTCTGCAATT TTCCAGATTA TTCAAAGTAT
 501 CAGGATGGGC ATGTGAAGTG ACTGACCTTA AGATGTTTCC ATTCTCCTGT
 551 GAATTTTAAC TTGAACTCAT TCCTGATGTT TGATACCCTG GTTGAAAACA
 601 ATTCAGTAAA GCATCCTGCC TCAGAATGAC TTTCCTATCA TGCTTCATGT
 651 GTCATTCCAA GGTTTCTTCA TGAGTCATTC CAAGTTTTCT AGTCCATACC
 701 ACAGTGCCTT GCAAAAAACA CCACATGAAT AAAGCAATAA AATTTGATTG
 751 TTAAGATACA GTAGTGGACC CTACTTATTC AGTCAATTAA GAGTAAGTTT
 801 TTTTATGTGG TTATTAAAAC AGTATGAACA ATTAGTCTAA CTCTGCATAG
 851 ACAGGGTCTA GATTTTGTTA ACCCAAATGT ATAACTGCAG TTAGCTTAAA
 901 TTACAATTTG AAGTCTTGTG GTTTTTATAT AGCTAGGCAC TTTATTACTC
 951 TTTTGAACTG AAAGCACACT CCCTTATAGG TTCATGTAAC TGTCCTGTAA
1001 TAAGGTGCTT ATAAATGGAA CAACTACACA GCCTAGTTTT GCCACAACCT
1051 TTAGCATCTA AAAAGTTTTA AAAGCTTCTA AATGTCTAAT ATAAAGGGAG
1101 ATGCTTATAG CCACAACATC TATTTTACCA ATATTGTTTC CATTACACTA
1151 CCTTGGATTT TGCATGAGTG AGTATAGTAA CCCAAGATGC CATAAAAAAA
1201 AACTTGATCG TTTTCTGACT TAATTAGTTA CTGTGGTTTC ACTAAAAGCT
1251 ACCGTGGTGG AGTGAAGTCA GTCAGGGAAG GTTTGTTTAT GTTACATTTA
1301 TTTCACCAGA ACTATTTTAA TATATCAAAG GGGTTTACTA TGCCAAACAA
1351 AATTCTAGGG AAAAATACTG CTAAAAATGG ATGCCTCATC AGAACATGCT
1401 GTTGAGTCCA ATGTGCCATA AGACATTTTA GCATGTTAAA TAGCACTTTT
1451 AATAGCAAAA AAAGGCACAT CAACTGCGAA GTTATCCTTA GTTTGCAAAT
1501 GCTTTTTCTA GATTAATGAT TTTTCAATCA TTAGGGTACT AGACACATCA
1551 GCCTAAAGTG GCATCTGGAA TTGAATGGAT TTACTGATAA TGATCAGTCT
1601 TTAGTCTTCC CTTTGTTATA TGACTTTATA GGTTATGATT GATCAAATTT
1651 ACGTTTTACT AATGGTAAGG GTGAGGGTCA TAGGGCAGGT TTTGGGTTTT
1701 CTAGTACTGT TGAAAACTGC AAGTATTGGC TATTTGTATA CTTAGCCATA
1751 ACTTGGTGAA AAAAAACCTG AGCAGTGTCT ATGTATTAAT GCGTTGGAAA
1801 GAAAGCTGCT TGTGTTTGCT TTGTTAATTG CCTCAGGATA TTTCTTTTAA
1851 AATAAGCTGT TTTAAGAGGA ACAGAAGGGA AATCTGCTAC CTAGTCTATA
1901 CACAGCGTGA ACCTCACAGG GGGCTTCTGA TACCCTCAAA CATGGAGAAC
1951 AGTAAGGGAG CAGAGTGGTT AAGGACTTTC AGGAACTTAA CTATTCTGGA
2001 ATAAGGAATG AATCAACTGA CCTTGGGCCA GCAGGTTTTT AACTAAATTG
2051 TTACTTGCCT TTCTCACCCA GTTAATCAGT CTCTGTACTT GTTTCCCTTT
2101 TTGAAACAAG TGTCTTGGTT AACTAATTCT GTTTTATGGT TGTGCTAAAT
2151 TCATAGCAGG TGCCTTATTC TTTGCTTTTA GTCAAACCAT TCCATATCAG
2201 AATTTTCCTT GGTTTACTAT AGATATTTGG CTTTAAGTTG TTGTTTGTGT
2251 TTTTTAATGT ACAATCTTCT GATAAATTTG ACTGTTAAAT TGCTATAGCT
2301 AGCAATCATT TTACATATGT AAAAAATTGC ATTCCCTTTG TATTTCATGT
2351 GTAATTCACC AATTAAGTGC AGTTTATATT CAGGTTGGAT TATGCATGTT
2401 TAGGTAAACG AAAGCTGTGT CTTACTTGAT TTATTCTTTA AAAATAAAGT
2451 TCCCTGAATA TTTGAAAAAA AAAAAAAAAA AAAAAAAA
```

BLAST Results

Entry HSCDN13 from database EMBL:
H.sapiens (TL5) mRNA from LNCaP cell line
Score = 1075, P = 5.8e-41, identities = 219/221

Entry AF100470_1 from database TREMBLNEW:
gene: "RAMP4"; product: "ribosome attached membrane protein 4"; Rattus norvegicus ribosome attached membrane protein 4 (RAMP4) mRNA, complete cds.
Score = 331, P = 3.9e-28, identities = 66/66, positives = 66/66, frame +1

Entry HSG19910 from database EMBL:
human STS A002B48.
Score = 530, P = 2.1e-17, identities = 108/109

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 316 bp to 513 bp; peptide length: 66
Category: strong similarity to known protein
Classification: Intacellular transport and traffic

```
 1 MVAKQRIRMA NEKHSKNITQ RGNVAKTSRN APEEKASVGP WLLALFIFVV
51 CGSAIFQIIQ SIRMGM
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_21116, frame 1

TREMBLNEW:RNO238236_1 gene: "ramp4"; product: "ribosome associated membrane protein RAMP4"; Rattus norvegicus mRNA for ribosome associated membrane protein RAMP4, N = 1, Score = 331, P = 6.2e-30

TREMBL:AF100470_1 gene: "RAMP4"; product: "ribosome attached membrane protein 4"; Rattus norvegicus ribosome attached membrane protein 4 (RAMP4) mRNA, complete cds., N = 1, Score = 331, P = 6.2e-30

>TREMBLNEW:RNO238236_1 gene: "ramp4"; product: "ribosome associated membrane protein RAMP4"; Rattus norvegicus mRNA for ribosome associated membrane protein RAMP4
Length = 75

HSPs:

Score = 331 (49.7 bits), Expect = 6.2e-30, P = 6.2e-30
Identities = 66/66 (100%), Positives = 66/66 (100%)

```
Query:    1 MVAKQRIRMANEKHSKNITQRGNVAKTSRNAPEEKASVGPWLLALFIFVVCGSAIFQIIQ 60
            MVAKQRIRMANEKHSKNITQRGNVAKTSRNAPEEKASVGPWLLALFIFVVCGSAIFQIIQ
Sbjct:   10 MVAKQRIRMANEKHSKNITQRGNVAKTSRNAPEEKASVGPWLLALFIFVVCGSAIFQIIQ 69

Query:   61 SIRMGM 66
            SIRMGM
Sbjct:   70 SIRMGM 75
```

No Pedant data available

DKFZphtes3_21n23
----

group: testes derived

DKFZphtes3_15j18 encodes a novel 148 amino acid protein with strong similarity to rat 7acomp protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific . genes.

strong similarity to rat 7acomp protein

on genomic level encoded by AF107885

Sequenced by LMU

Locus: /map="14q24.3"

Insert length: 3122 bp
Poly A stretch at pos. 3070, polyadenylation signal at pos. 3045

```
   1 GGAAAACCTC GTGGGCTCAG CCCGGGAGAA AGGGCCAGGG AAGTTGGGTG
  51 GTTCTGTGCT TGGTCTGTCA ATGGAGGAGA TCAAAGTTTT ACGAAGGGTG
 101 AAGGAGGAGA ATGATCGGCG AGGTGGATTT ATTCGCATAT TTCCTACATC
 151 TGAGACATGG GAAATATATG GGTCCTACCT CGAGCATAAG ACCTCAATGA
 201 ACTATATGCT GGCAACACGC CTCTTCCAGG ACAGGGGAAA CCCAAGAAGA
 251 AGCTTATTGA CAGGAAGAAC ACGAATGACT GCTGATGGAG CGCCAGAATT
 301 GAAGATAGAG AGTCTGAATT CAAAGGCCAA GCTGCATGCT GCACTTTACG
 351 AGAGGAAGCT CCTGTCTCTG GAGGTGCGAA AACGTAGACG ACGGAGTAGC
 401 AGATTGAGGG CAATGAGGCC AAAATACCCA GTGATTACCC AACCAGCTGA
 451 AATGAATGTT AAAACTGAGA CAGAGAGTGA AGAGGAGGAA GAAGTCGCAT
 501 TAGATAATGA AGATGAAGAA CAGGAGGCTT CCCAGGAGGA GTCTGCAGGA
 551 TTTCTTAGAG AAAATCAAGC CAAATATACA CCCTCATTGA CAGCTTTGGT
 601 AGAAAATACA CCCAAAGAAA ATTCCATGAA AGTTCGTGAA TGGAATAATA
 651 AAGGTGGACA CTGCTGCAAA CTTGAGACTC AGGAGCTAGA GCCTAAATTT
 701 AACCTGATGC AGATTCTTCA AGATAATGGC AATCTTAGCA AAATGCAGGC
 751 CCGAATAGCA TTCTCTGCCT ATCTCCAGCA TGTTCAAATT CGCCTGATGA
 801 AAGACAGTGG CGGTCAGACG TTCAGTGCCA GTTGGGCTGC CAAAGAGGAT
 851 GAACAGATGG AGCTGGTTGT TCGTTTCCTC AAGCGAGCAT CAAATAACCT
 901 CCAGCATTCA CTGAGGATGG TATTACCCAG TCGACGATTG GCACTTCTGG
 951 AACGCAGAAG AATCCTGGCC CACCAGCTGG GTGACTTTAT CATTGTATAC
1001 AACAAGGAAA CAGAACAAAT GGCTGAAAAG AAATCAAAGA AGAAAGTTGA
1051 GGAAGAAGAG GAAGATGGGG TGAATATGGA AAACTTTCAG GAGTTCATCA
1101 GACAAGCAAG TGAGGCTGAA CTGGAGGAGG TGTTGACTTT TTATACCCAA
1151 AAGAACAAGT CTGCTAGTGT CTTCCTGGGG ACTCACTCTA AAATTTCTAA
1201 GAACAACAAC AATTATTCTG ATAGTGGGGC AAAAGGTGAT CACCCTGAGA
1251 CTATAATGGA AGAAGTGAAA ATAAAGCCAC CTAAACAGCA ACAGACGACA
1301 GAAATTCATT CTGATAAATT ATCTCGATTT ACCACTTCAG CAGAAAAAGA
1351 GGCAAAAATTA GTTTATAGCA ATTCCTCCTC TGGTCCTACT GCTACTCTGC
1401 AGAAAATTCC CAACACCCAT TTGTCATCTG TTACAACCTC TGACCTCTCT
1451 CCAGGGCCTT GCCACCATTC TTCTTTATCT CAAATTCCTT CAGCTATCCC
1501 CAGCATGCCT CACCAGCCAA CAATTTTACT GAACACAGTC TCTGCCAGTG
1551 CTTCTCCCTG CCTACATCCC GGGGCACAGA ACATCCCAAG CCCTACTGGC
1601 CTGCCACGCT GTCGATCAGG AAGTCACACC ATTGGTCCCT TTTCTTCCTT
1651 CCAAAGTGCT GCACACATCT ATAGCCAGAA ACTGTCTCGT CCCTCTTCAG
1701 CAAAGGCAGG ATCGTGCTAT CTAAACAAGC ATCATTCAGG AATAGCCAAA
1751 ACACAAAAAG AGGGAGAAGA TGCTTCTTTA TATAGCAAAC GGTACAACCA
1801 AAGTATGGTT ACAGCTGAAC TTCAGCGGCT AGCTGAGAAG CAGGCAGCGA
1851 GACAGTATTC TCCATCCAGC CACATCAACC TCCTCACCCA ACAGGTAACA
1901 AACCTGAATT TGGCAACTGG CATCATAAAC AGAAGCAGTG CTTCAGCTCC
1951 CCCAACCCTC CGACCCATCA TCAGTCCTAG TGGCCCGACA TGGTCTACAC
2001 AGTCAGACCC CCAAGCTCCC GAGAATCACT CCAGCTCTCC TGGAAGCAGG
2051 AGCCTGCAGA CAGGGGGATT TGCCTGGGAA GGAGAAGTAG AAAACAACGT
2101 GTACAGCCAG GCTACAGGGG TGGTCCCCCA GCACAAGTAT CACCCCACAG
2151 CAGGCAGCTA TCAGCTTCAA TTTGCCCTGC AGCAACTTGA ACAACAAAAA
2201 CTTCAGTCCC GGCAGCTCCT GGACCAGAGT CGAGCCCGGC ACCAGGCAAT
2251 CTTTGGCAGC CAGACACTAC CTAACTCCAA TTTATGGACA ATGAATAATG
2301 GTGCAGGTTG TAGAATTTCC AGTGCCACAG CTAGTGGCCA GAAGCCAACC
2351 ACTCTGCCAC AAAAAGTGGT ACCACCTCCA AGTTCTTGCG CCTCCCTGGT
2401 TCCCAAACCC CCACCCAACC ACGAACAAGT GCTCAGAAGG GCAACATCCC
2451 AGAAAGCTTC CAATACCCGC TTCAGATCCT CCTTTCAAAA CTATTTGTGG
2501 TATTTCTTCC AAGCAGTCAG CTGAACTGAG GACGACAGCC TACAAACAAC
2551 TACATGCATC TGAACTGTCT CTTGTAAATG AGCTTTTTTC AGAGCCAGAA
2601 TCATACTCTC CAGGAAATAT GGAGAAAGAA ACCTGAGGAG ATTGAAGTTT
2651 GCCAGGCACA AGGGCAAAAC TCAGACTGAA TGAATTTGAA AGGGTGGGGC
2701 CAAAGATGTT GTAACCTGGG AGACTTCTCT GAAGAAAGAA AACTGTTTAA
```

```
2751 GAAACACAGA CTGAACTGCA GTACTTTTCC TTAAATAGCT GAGATGACCT
2801 TCTTTACCCT GGGCTTAGGT GATTCTCATC AGGGTGACCT GAGTGGAAGT
2851 TGGTGGTAAC GACTGTTCTG TGTCAGCACC CAGGACAGTG GTGTCTGTTA
2901 AGGCTGCCAG GGATTAGCAG GGAGGAAAGC CATCAGGACT GGGTAGCCTG
2951 GTAGCACCAA ATCCCAATTA ATGTTACCTG AACATGTGGT GAGGTCAGCC
3001 GTATGATGAA AGATGTTTAA GAGATTAATG TCAGAAGAAT ATGAAAATAA
3051 ACACCGGCTT AAAAAATGTT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3101 AAAAAAAAAA AAAAAAAAAA AA
```

BLAST Results

Entry AF107885 from database EMBL:
Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes.
Score = 3042, P = 3.0e-219, identities = 610/612
5 exons matching 1893-3070

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 71 bp to 2521 bp; peptide length: 817
Category: strong similarity to known protein

```
  1 MEEIKVLRRV KEENDRRGGF IRIFPTSETW EIYGSYLEHK TSMNYMLATR
 51 LFQDRGNPRR SLLTGRTRMT ADGAPELKIE SLNSKAKLHA ALYERKLLSL
101 EVRKRRRRSS RLRAMRPKYP VITQPAEMNV KTETESEEEE EVALDNEDEE
151 QEASQEESAG FLRENQAKYT PSLTALVENT PKENSMKVRE WNNKGGHCCK
201 LETQELEPKF NLMQILQDNG NLSKMQARIA FSAYLQHVQI RLMKDSGGQT
251 FSASWAAKED EQMELVVRFL KRASNNLQHS LRMVLPSRRL ALLERRRILA
301 HQLGDFIIVY NKETEQMAEK KSKKKVEEEE EDGVNMENFQ EFIRQASEAE
351 LEEVLTFYTQ KNKSASVFLG THSKISKNNN NYSDSGAKGD HPETIMEEVK
401 IKPPKQQQTT EIHSDKLSRF TTSAEKEAKL VYSNSSSGPT ATLQKIPNTH
451 LSSVTTSDLS PGPCHHSSLS QIPSAIPSMP HQPTILLNTV SASASPCLHP
501 GAQNIPSPTG LPRCRSGSHT IGPFSSFQSA AHIYSQKLSR PSSAKAGSCY
551 LNKHHSGIAK TQKEGEDASL YSKRYNQSMV TAELQRLAEK QAARQYSPSS
601 HINLLTQQVT NLNLATGIIN RSSASAPPTL RPIISPSGPT WSTQSDPQAP
651 ENHSSSPGSR SLQTGGFAWE GEVENNVYSQ ATGVVPQHKY HPTAGSYQLQ
701 FALQQLEQQK LQSRQLLDQS RARHQAIFGS QTLPNSNLWT MNNGAGCRIS
751 SATASGQKPT TLPQKVVPPP SSCASLVPKP PPNHEQVLRR ATSQKASNTR
801 FRSSFQNYLW YFFQAVS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_21n23, frame 2

TREMBL:AF064856_1 product: "7acomp protein"; Rattus sp. 7acomp protein mRNA, complete cds., N = 1, Score = 1845, P = 2.2e-190

TREMBL:AF107885_3 product: "unknown"; Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes., N = 1, Score = 443, P = 5.3e-41

TREMBL:AF107885_4 product: "unknown"; Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes., N = 1, Score = 265, P = 8.2e-22

>TREMBL:AF064856_1 product: "7acomp protein"; Rattus sp. 7acomp protein
    mRNA, complete cds.
        Length = 436

HSPs:

Score = 1845 (276.8 bits), Expect = 2.2e-190, P = 2.2e-190
Identities = 369/435 (84%), Positives = 395/435 (90%)

```
Query:   115 MRPKYPVITQPAEMNVKTETESEEEEEVALDNEDEEQEASQEESAGFLRENQAKYTPSLT 174
             MRPKYPVIT PAEMN+KTETESEEEEEV LDNEDEEQEASQEESAG L ENQAKYTPSLT
Sbjct:     1 MRPKYPVITLPAEMNIKTETESEEEEEVGLDNEDEEQEASQEESAGSLAENQAKYTPSLT 60

Query:   175 ALVENTPKENSMKVREWNNKGGHCCKLETQELEPKFNLMQILQDNGNLSKMQARIAFSAY 234
              +VEN+P+EN+MKV EW NKG  CCK+ETQE E KFNLMQILQDNGNLSK+QAR+AFSAY
Sbjct:    61 VIVENSPRENAMKVAEWTNKGESCCKIETQEPESKFNLMQILQDNGNLSKVQARLAFSAY 120

Query:   235 LQHVQIRLMKDSGGQTFSASWAAKEDEQMELVVRFLKRASNNLQHSLRMVLPSRRLALLE 294
             LQHVQ+RL KDSGGQT S SWAAKEDEQMELVVRFLKRAS+NLQHSLRMVLPSRRLALLE
Sbjct:   121 LQHVQVRLTKDSGGQTLSPSWAAKEDEQMELVVRFLKRASSNLQHSLRMVLPSRRLALLE 180

Query:   295 RRRILAHQLGDFIIVYNKETEQMAEKKSKKKVEEEEEDGVNMENFQEFIRQASEAELEEV 354
             RRRILAHQLGDFI+VYNKETEQMAEKKSKKK+EEEEEDGVN E+FQEFIRQASEAELEEV
Sbjct:   181 RRRILAHQLGDFIVVYNKETEQMAEKKSKKKLEEEEEDGVNAESFQEFIRQASEAELEEV 240

Query:   355 LTFYTQKNKSASVFLGTHSKISKNNNNYSDSGAKGDHPETIMEEVKIKPPKQQQTTEIHS 414
             LTFYTQKNKSASVFLGTHSK SKN+++YSDSGAKGDHPETI +EVKIK PKQQQ TEIHS
Sbjct:   241 LTFYTQKNKSASVFLGTHSKSSKNSSSYSDSGAKGDHPETI-QEVKIKQPKQQQATEIHS 299

Query:   415 DKLSRFTTSAEKEAKLVYSNSSS--GPTATL-QKIPNTHLSSV-TTSDLSPGPCHHSSLS 470
             DKLSRFTTSA KEAKLVY+N SS  GP A L Q++P+THLSS+ TTS LS GP HHSSLS
Sbjct:   300 DKLSRFTTSAGKEAKLVYTNCSSFSGPAAVLLQRLPSTHLSSIITTSTLSSGPGHHSSLS 359

Query:   471 QIPSAIPSMPHQPTILLNTVSASASPCLHPGAQNIPSPTGLPRCRSGSHTIGPFSSFQSA 530
             QI  AIPSMPHQ  +LLN V  SASP +HPG  N+ SP GLPRCRSGS+TIGPFSSFQSA
Sbjct:   360 QISPAIPSMPHQSALLLNPVPDSASPPVHPGTPNV-SPAGLPRCRSGSYTIGPFSSFQSA 418

Query:   531 AHIYSQKLSRPSSAKAG 547
             AHIYSQKLSRPSSAKAG
Sbjct:   419 AHIYSQKLSRPSSAKAG 435
```

Pedant information for DKFZphtes3_21n23, frame 2

Report for DKFZphtes3_21n23.2

```
[LENGTH]      817
[MW]          91522.09
[pI]          9.32
[HOMOL]       TREMBL:AF064856_1 product: "7acomp protein"; Rattus sp. 7acomp protein mRNA,
complete cds. 1e-166
[PROSITE]     MYRISTYL          6
[PROSITE]     CAMP_PHOSPHO_SITE     4
[PROSITE]     CK2_PHOSPHO_SITE      12
[PROSITE]     TYR_PHOSPHO_SITE      1
[PROSITE]     PKC_PHOSPHO_SITE      15
[PROSITE]     ASN_GLYCOSYLATION     7
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY    13.83 %

SEQ   MEEIKVLRRVKEENDRRGGFIRIFPTSETWEIYGSYLEHKTSMNYMLATRLFQDRGNPRR
SEG   ............................................................
PRD   ccchhhhhhhhhhhcccccceeeecccccceeeeecceeeecccchhhhhhhhhhcccccc

SEQ   SLLTGRTRMTADGAPELKIESLNSKAKLHAALYERKLLSLEVRKRRRRSSRLRAMRPKYP
SEG   ................................xxxxxxxxxxxxxxxxxxx.......
PRD   cccccccceeecccccceeeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccc

SEQ   VITQPAEMNVKTETESEEEEEVALDNEDEEQEASQEESAGFLRENQAKYTPSLTALVENT
SEG   ......xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....................
PRD   ceeecchhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhcccccceeeeeccc

SEQ   PKENSMKVREWNNKGGHCCKLETQELEPKFNLMQILQDNGNLSKMQARIAFSAYLQHVQI
SEG   ............................................................
PRD   ccccccceeeeecccccccchhhhhhhcccchhhhhhhccccchhhhhhhhhhhhhhhhh

SEQ   RLMKDSGGQTFSASWAAKEDEQMELVVRFLKRASNNLQHSLRMVLPSRRLALLERRRILA
SEG   ..........................................xxxxxxxxxxxxxx.
PRD   hhhhccccccceeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhh

SEQ   HQLGDFIIVYNKETEQMAEKKSKKKVEEEEEDGVNMENFQEFIRQASEAELEEVLTFYTQ
SEG   ..................xxxxxxxxxxxxx.............................
PRD   hhccceeeeeehhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhh

SEQ   KNKSASVFLGTHSKISKNNNNYSDSGAKGDHPETIMEEVKIKPPKQQQTTEIHSDKLSRF
SEG   ............................................................
PRD   cccccceeeccccccccccccccccccccchhhhhhhccccccceeeeeccccccccc
```

```
SEQ    TTSAEKEAKLVYSNSSSGPTATLQKIPNTHLSSVTTSDLSPGPCHHSSLSQIPSAIPSMP
SEG    ............................................................
PRD    hhhhhhheeeecccccccceeeeccccccccccccccccccccccccccccccccccccc

SEQ    HQPTILLNTVSASASPCLHPGAQNIPSPTGLPRCRSGSHTIGPFSSFQSAAHIYSQKLSR
SEG    ............................................................
PRD    cccceeeeccccccccccccccccccccccccccccccccccccchhhhhhhhhhccc

SEQ    PSSAKAGSCYLNKHHSGIAKTQKEGEDASLYSKRYNQSMVTAELQRLAEKQAARQYSPSS
SEG    ............................................................
PRD    cccccccceeeeccccccccccccccceeeecchhhhhhhhhhhhhhhhhhhhhhcccc

SEQ    HINLLTQQVTNLNLATGIINRSSASAPPTLRPIISPSGPTWSTQSDPQAPENHSSSPGSR
SEG    ..xxxxxxxxxxxxx.............................................
PRD    cccccccccccccccccccccccccccccceeeecccccccccccccccccccccccccc

SEQ    SLQTGGFAWEGEVENNVYSQATGVVPQHKYHPTAGSYQLQFALQQLEQQKLQSRQLLDQS
SEG    ....................................xxxxxxxxxxxxxxxxxxxxx...
PRD    cccccccceeeeeccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhh

SEQ    RARHQAIFGSQTLPNSNLWTMNNGAGCRISSATASGQKPTTLPQKVVPPPSSCASLVPKP
SEG    ............................................................
PRD    hhhhhhhcccccccccceeeecccccceeeeeeeccccccccccccceeeccccccceeecccc

SEQ    PPNHEQVLRRATSQKASNTRFRSSFQNYLWYFFQAVS
SEG    .....................................
PRD    cccchhhhhhhhhhhcccccccccccceeeeeeeccc
```

Prosite for DKFZphtes3_21n23.2

PS00001	221->225	ASN_GLYCOSYLATION	PDOC00001
PS00001	362->366	ASN_GLYCOSYLATION	PDOC00001
PS00001	381->385	ASN_GLYCOSYLATION	PDOC00001
PS00001	434->438	ASN_GLYCOSYLATION	PDOC00001
PS00001	576->580	ASN_GLYCOSYLATION	PDOC00001
PS00001	620->624	ASN_GLYCOSYLATION	PDOC00001
PS00001	652->656	ASN_GLYCOSYLATION	PDOC00001
PS00004	106->110	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	107->111	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	271->275	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	789->793	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	64->67	PKC_PHOSPHO_SITE	PDOC00005
PS00005	109->112	PKC_PHOSPHO_SITE	PDOC00005
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00005	185->188	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	287->290	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	359->362	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	543->546	PKC_PHOSPHO_SITE	PDOC00005
PS00005	561->564	PKC_PHOSPHO_SITE	PDOC00005
PS00005	572->575	PKC_PHOSPHO_SITE	PDOC00005
PS00005	629->632	PKC_PHOSPHO_SITE	PDOC00005
PS00005	793->796	PKC_PHOSPHO_SITE	PDOC00005
PS00006	35->39	CK2_PHOSPHO_SITE	PDOC00006
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	134->138	CK2_PHOSPHO_SITE	PDOC00006
PS00006	136->140	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	347->351	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	422->426	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	561->565	CK2_PHOSPHO_SITE	PDOC00006
PS00006	643->647	CK2_PHOSPHO_SITE	PDOC00006
PS00007	563->572	TYR_PHOSPHO_SITE	PDOC00007
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	248->254	MYRISTYL	PDOC00008
PS00008	510->516	MYRISTYL	PDOC00008
PS00008	557->563	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00008	756->762	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_21n23.2)

DKFZphtes3_22c23
----

group: testes derived

DKFZphtes3_22c23 encodes a novel 223 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, 3 EST hits (two from a testis library)

Sequenced by LMU

Locus: /map="9q34"

Insert length: 1113 bp
Poly A stretch at pos. 1073, polyadenylation signal at pos. 1055

```
   1 GGTGGGCAAA GGCATCTTCC TCTGGGAAGG ACTGGCACAA GCACTTGGTC
  51 CCTGGGTTGT GTGCCTGGGA GGCCGGGATC AGGGCTGGCC CTCTTTCTCC
 101 CTGGCAAAGC AAAACCTCCC TTTTACTACT ATCAAGGGGA AGTAACTTGA
 151 AGGTGCCTGT GGCAGGCAGC ACCTTGAGCC AACAGGAACC ATTGACATGC
 201 GAGGCCCAGG GCAGGCAGAC TGTGCAGTGG CCATTGGGCG GCCCCTCGGG
 251 GAGGTGGTGA CCCTCCGCGT CCTTGAGAGT TCTCTCAACT GCAGTGCGGG
 301 GGACATGTTG CTGCTTTGGG GCCGGCTCAC CTGGAGGAAG ATGTGCAGGA
 351 AGCTGTTGGA CATGACTTTC AGCTCCAAGA CCAACACGCT GGTGGTGAGG
 401 CAGCGCTGCG GGCGGCCAGG AGGTGGGGTG CTGCTGCGGT ATGGGAGCCA
 451 GCTTGCTCCT GAAACCTTCT ACAGAGAATG TGACATGCAG CTCTTTGGGC
 501 CCTGGGGTGA AATCGTGAGC CCCTCGCTGA GTCCAGCCAC GAGTAATGCA
 551 GGGGGCTGCC GGCTCTTCAT TAATGTGGCT CCGCACGCAC GGATTGCCAT
 601 CCATGCCCTG GCCACCAACA TGGGCGCTGG GACCGAGGGA GCCAATGCCA
 651 GCTACATCTT GATCCGGGAC ACCCACAGCT TGAGGACCAC AGCGTTCCAT
 701 GGGCAGCAGG TGCTCTACTG GGAGTCAGAG AGCAGCCAGG CTGAGATGGA
 751 GTTCAGCGAG GGCTTCCTGA AGGCTCAGGC CAGCCTGCGG GGCCAGTACT
 801 GGACCCTCCA ATCATGGGTA CCGGAGATGC AGGACCCTCA GTCCTGGAAG
 851 GGAAAGGAAG GAACCTGAGG GTCATTGAAC ATTTGTTCCG TGTCTGGCCA
 901 GCCCTGGAGG GTTGACCCCT GGTCTCAGTG CTTTCCAATT CGAACTTTTT
 951 CCAATCTTAG GTATCTACTT TAGAGTCTTC TCCAATGTCC AAAAGGCTAG
1001 GGGGTTGGAG GTGGGGACTC TGGAAAAGCA GCCCCCATTT CCTCGGGTAC
1051 CAATAAATAA AACATGCAGG CTGAAAAAAA AAAAAAAAAA AAAAAAAAAA
1101 AAAAAAAAAA AAA
```

BLAST Results

Entry HSAC1644 from database EMBL:
Genomic sequence from Human 9q34, complete sequence.
Score = 2072, P = 8.8e-225, identities = 422/430
5 exons Bp 41969-38232

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 197 bp to 865 bp; peptide length: 223
Category: putative protein

```
   1 MRGPGQADCA VAIGRPLGEV VTLRVLESSL NCSAGDMLLL WGRLTWRKMC
  51 RKLLDMTFSS KTNTLVVRQR CGRPGGVLL  RYGSQLAPET FYRECDMQLF
 101 GPWGEIVSPS LSPATSNAGG CRLFINVAPH ARIAIHALAT NMGAGTEGAN
 151 ASYILIRDTH SLRTTAFHGQ QVLYWESESS QAEMEFSEGF LKAQASLRGQ
 201 YWTLQSWVPE MQDPQSWKGK EGT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_22c23, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_22c23, frame 2

Report for DKFZphtes3_22c23.2

[LENGTH]	223	
[MW]	24546.19	
[pI]	8.57	
[PROSITE]	MYRISTYL	4
[PROSITE]	CK2_PHOSPHO_SITE	2
[PROSITE]	PKC_PHOSPHO_SITE	6
[PROSITE]	ASN_GLYCOSYLATION	2
[KW]	Alpha_Beta	

```
SEQ  MRGPGQADCAVAIGRPLGEVVTLRVLESSLNCSAGDMLLLWGRLTWRKMCRKLLDMTFSS
PRD  ccccccccceeeeccccccceeeeehhhhhcccccchhhhhhchhhhhhhhhhhhhhhccc

SEQ  KTNTLVVRQRCGRPGGGVLLRYGSQLAPETFYRECDMQLFGPWGEIVSPSLSPATSNAGG
PRD  ccceeeeeecccccccceeeecccccchhhhhhhhcccccceeeecccccccccccc

SEQ  CRLFINVAPHARIAIHALATNMGAGTEGANASYILIRDTHSLRTTAFHGQQVLYWESESS
PRD  ceeeeeeccccceeehhhhhhhhccccccccceeeeeeccccceeeccccceeeeecccccc

SEQ  QAEMEFSEGFLKAQASLRGQYWTLQSWVPEMQDPQSWKGKEGT
PRD  hhhhhhhcchhhhhhhhhhcccccccccccccccccccccccc
```

Prosite for DKFZphtes3_22c23.2

PS00001	31->35	ASN_GLYCOSYLATION	PDOC00001
PS00001	150->154	ASN_GLYCOSYLATION	PDOC00001
PS00005	22->25	PKC_PHOSPHO_SITE	PDOC00005
PS00005	45->48	PKC_PHOSPHO_SITE	PDOC00005
PS00005	59->62	PKC_PHOSPHO_SITE	PDOC00005
PS00005	161->164	PKC_PHOSPHO_SITE	PDOC00005
PS00005	196->199	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00006	33->37	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00008	5->11	MYRISTYL	PDOC00008
PS00008	145->151	MYRISTYL	PDOC00008
PS00008	148->154	MYRISTYL	PDOC00008
PS00008	199->205	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_22c23.2)

DKFZphtes3_22g2
----

group: nucleic acid management

DKFZphtes3_22g2encodes a novel 1230 amino acid protein with nearly identical to rat TIP120.

TATA-binding protein TBP is a central component for transcriptional regulation and is a target for various transcription regulators. TBP-interacting protein 120 (TIP120) is a protein interacting with the TATA-binding protein (TBP). The novel protein is the human ortholog of rat TIP120. The novel TBP-binding protein is considered to participate in transcription regulation through the interaction with TBP.

The new protein can find application in modulation of gene transcription.

KIAA0829, complete cds, nearly identical to rat TIP120

complete cDNA, complete cds, EST hits,

Sequenced by LMU

Locus: /map="387.3 cR from top of Chr12 linkage group"

Insert length: 5387 bp
Poly A stretch at pos. 5352, polyadenylation signal at pos. 5335

```
   1 GGGAGCGAGT GCGGAGCGAG TGGGAGCGAG ACGGCCCTGA GTGGAAGTGT
  51 CTGGCTCCCC GTAGAGGCCC TTCTGTACGC CCCGCCGCCC ATGAGCTCGT
 101 TCTCACGCGA ACAGCGCCGT CGTTAGGCTG GCTCTGTAGC CTCGGCTTAC
 151 CCCGGGACAG GCCCACGCCT CGCCAGGGAG GGGGCAGCCC GTCGAGGCGC
 201 CTCCCTAGTC AGCGTCGGCG TCGCGCTGCG ACCCTGGAAG CGGGAGCCGC
 251 CGCGAGCGAG AGGAGGAGCT CCAGTGGCGG CGGCGGCGGC GGCAGCGGCA
 301 GCGGGCAGCA GCTCCAGCAG CGCCAGCAGG CGGGATCGAG GCCGTCAACA
 351 TGGCGAGCGC CTCGTACCAC ATTTCCAATT TGCTGGAAAA AATGACATCC
 401 AGCGACAAGG ACTTTAGGTT TATGGCTACA AATGATTTGA TGACGGAACT
 451 GCAGAAAGAT TCCATCAAGT TGGATGATGA TAGTGAAAGG AAAGTAGTGA
 501 AAATGATTTT GAAGTTATTG GAAGATAAAA ATGGAGAGGT ACAGAATTTA
 551 GCTGTCAAAT GTCTTGGTCC TTTAGTGAGT AAAGTGAAAG AATACCAAGT
 601 AGAGACAATT GTAGATACCC TCTGCACTAA CATGCTTTCT GATAAAGAAC
 651 AACTTCGAGA CATTTCAAGT ATTGGTCTTA AAACAGTAAT TGGAGAACTT
 701 CCTCCAGCTT CCAGTGGCTC TGCATTAGCT GCTAATGTAT GTAAAAAGAT
 751 TACTGGACGT CTTACAAGTG CAATAGCAAA ACAGGAAGAT GTCTCTGTTC
 801 AGCTAGAAGC CTTGGATATT ATGGCTGATA TGTTGAGCAG GCAAGGAGGA
 851 CTTCTTGTTA ATTTCCATCC TTCAATTCTG ACCTGTCTAC TTCCCCAGTT
 901 GACCAGCCCT AGACTTGCAG TGAGGAAAAG AACCATTATC GCTCTTGGCC
 951 ATCTGGTTAT GAGCTGTGGA AATATAGTTT TTGTAGATCT TATTGAACAT
1001 CTGTTGTCAG AGTTGTCCAA AAATGATTCT ATGTCAACAA CAAGAACCTA
1051 CATACAATGT ATTGCTGCTA TTAGTAGGCA AGCTGGTCAT AGAATAGGTG
1101 AATACCTTGA GAAGATAATT CCTTTGGTGG TAAAATTTTG CAATGTAGAT
1151 GATGATGAAT TAAGAGAGTA CTGTATTCAA GCCTTTGAAT CATTTGTAAG
1201 AAGATGTCCT AAGGAAGTAT ATCCTCATGT TTCTACCATT ATAAATATTT
1251 GTCTTAAATA TCTTACCTAT GATCCAAATT ATAATTACGA TGATGAAGAT
1301 GAAGATGAAA ATGCAATGGA TGCTGATGGT GGTGATGATG ATGATCAAGG
1351 GAGTGATGAT GAATACAGTG ATGATGATGA CATGAGTTGG AAAGTGAGAC
1401 GTGCAGCTGC GAAGTGCTTG GATGCTGTAG TTAGCACAAG GCATGAAATG
1451 CTTCCAGAAT TCTACAAGAC CGTCTCTCCT GCACTAATAT CCAGATTTAA
1501 AGAGCGTGAA GAGAATGTAA AGGCAGATGT TTTTCACGCA TACCTTTCTC
1551 TTTTGAAGCA AACTCGTCCT GTACAAAGTT GGCTATGTGA CCCTGATGCA
1601 ATGGAGCAGG GAGAAACACC TTTAACAATG CTTCAGAGTC AGGTTCCCAA
1651 CATTGTTAAA GCTCTTCACA AACAGATGAA AGAAAAAAGT GTGAAGACCC
1701 GACAGTGTTG TTTTAACATG TTAACTGAGC TGGTAAATGT ATTACCTGGG
1751 GCCCTAACTC AACACATTCC TGTACTTGTA CCAGGAATCA TTTTCTCACT
1801 GAATGATAAA TCAAGCTCAT CGAATTTGAA GATCGATGCT TTGTCATGTC
1851 TATACGTAAT CCTCTGTAAC CATTCTCCTC AAGTCTTCCA TCCTCACGTT
1901 CAGGCTTTGG TTCCTCCAGT GGTGGCTTGT GTTGGAGACC CATTTTACAA
1951 AATTACATCT GAAGCACTTC TTGTTACTCA ACAGCTTGTC AAAGTAATTC
2001 GTCCTTTAGA TCAGCCTTCC TCGTTTGATG CAACTCCTTA TATCAAAGAT
2051 CTATTTACCT GTACCATTAA GAGATTAAAA GCAGCTGACA TTGATCAGGA
2101 AGTCAAGGAA AGGGCTATTT CCTGTATGGG ACAAATTATT TGCAACCTTG
2151 GAGACAATTT GGGTTCTGAC TTGCCTAATA CACTTCAGAT TTTCTTGGAG
2201 AGACTAAAGA ATGAAATTAC CAGGTTAACT ACAGTAAAGG CATTGACACT
2251 GATTGCTGGG TCACCTTTGA AGATAGATTT GAGGCCTGTT CTGGGAGAAG
2301 GGGTTCCTAT CCTTGCTTCA TTTCTTAGAA AAAACCAGAG AGCTTTGAAA
2351 CTGGGTACTC TTTCTGCCCT TGATATTCTA ATAAAAAACT ATAGTGACAG
2401 CTTGACAGCT GCCATGATTG ATGCAGTTCT AGATGAGCTC CCACCTCTTA
2451 TCAGCGAAAG TGATATGCAT GTTTCACAAA TGGCCATCAG TTTTCTTACC
2501 ACTTTGGCAA AAGTATATCC CTCCTCCCTT TCAAAGATAA GTGGATCCAT
2551 TCTCAATGAA CTTATTGGAC TTGTGAGATC ACCCTTATTG CAGGGGGGAG
```

```
2601 CTCTTAGTGC CATGCTAGAC TTTTTCCAAG CTCTGGTTGT CACTGGAACA
2651 AATAATTTAG GATACATGGA TTTGTTGCGC ATGCTGACTG GTCCAGTTTA
2701 CTCTCAGAGC ACAGCTCTTA CTCATAAGCA GTCTTATTAT TCCATTGCCA
2751 AATGTGTAGC TGCCCTTACT CGAGCATGCC CTAAAGAGGG ACCAGCTGTA
2801 GTAGGTCAGT TTATTCAAGA TGTCAAGAAC TCAAGGTCTA CAGATTCCAT
2851 TCGTCTCTTA GCTCTACTTT CTCTTGGAGA AGTTGGGCAT CATATTGACT
2901 TAAGTGGACA GTTGGAACTA AAATCTGTAA TACTAGAAGC TTTCTCATCT
2951 CCTAGTGAAG AAGTCAAATC AGCTGCATCC TATGCATTAG GCAGCATTAG
3001 TGTGGGCAAC CTTCCTGAAT ATCTGCCGTT TGTCCTGCAA GAAATAACTA
3051 GTCAACCCAA AAGGCAGTAT CTTTTACTTC ATTCCTTGAA GGAAATTATT
3101 AGCTCTGCAT CAGTGGTGGG CCTTAAACCA TATGTTGAAA ACATCTGGGC
3151 CTTATTACTA AAGCACTGTG AGTGTGCAGA GGAAGGAACC AGAAATGTTG
3201 TTGCTGAATG TCTAGGAAAA CTCACTCTAA TTGATCCAGA AACTCTCCTT
3251 CCACGGCTTA AGGGGTACTT GATATCAGGC TCATCATATG CCCGAAGCTC
3301 AGTGGTTACG GCTGTGAAAT TTACAATTTC TGACCATCCA CAACCTATTG
3351 ATCCACTGTT AAAGAACTGC ATAGGTGATT TCCTAAAAAC TTTGGAAGAC
3401 CCAGATTTGA ATGTGAGAAG AGTAGCCTTG GTCACATTTA ATTCAGCAGC
3451 ACATAACAAG CCATCATTAA TAAGGGATCT ATTGGATACT GTTCTTCCAC
3501 ATCTTTACAA TGAAACAAAA GTTAGAAAGG AGCTTATAAG AGAGGTAGAA
3551 ATGGGTCCAT TTAAACATAC GGTTGATGAT GGTCTGGATA TTAGAAAGGC
3601 AGCATTTGAG TGTATGTACA CACTTCTAGA CAGTTGTCTT GATAGACTTG
3651 ATATCTTTGA ATTTCTAAAT CATGTTGAAG ATGGTTTGAA GGACCATTAT
3701 GATATTAAGA TGCTGACATT TTTAATGTTG GTGAGACTGT CTACCCTTTG
3751 TCCAAGTGCA GTACTGCAGA GGTTGGACCG ACTTGTTGAG CCATTACGTG
3801 CAACATGTAC AACTAAGGTA AAGGCAAACT CAGTAAAGCA GGAGTTTGAA
3851 AAACAAGATG AATTAAAGCG ATCTGCCATG AGAGCAGTAG CAGCACTGCT
3901 AACCATTCCA GAAGCAGAGA AGAGTCCACT GATGAGTGAA TTCCAGTCAC
3951 AGATCAGTTC TAACCCTGAG CTGGCGGCTA TCTTTGAAAG TATCCAGAAA
4001 GATTCATCAT CTACTAACTT GGAATCAATG GACACTAGTT AGATGTTTGT
4051 TCACCATGGG GACCATTACA TATGACCATA CAATGCACTG AATTGACAGG
4101 TTAATCATAA GACATGGAAA GAGAAGTGTC TAAAAGCTTC AAAATGTTCC
4151 ACTTTTTTTT CCTTCATGGA GACTGTTTGT TTGGCTTTCT TCCATTGTTG
4201 TTTTTGTAGC ATTTATTTCA GAAATGTGTA TTTCCATAAT CCAGAGGTTG
4251 TAAAACCACT AGTGTTTTAG TGGTTACAGC AACATTTGAA ATGGAAACTA
4301 AAAGTTAGGA TTTTATGGAG TATGGAGATA GGGTCCAGTA TCTATTTACC
4351 CTGTAATGTT TAGGATTAAA ATGTTAAAAT TTTGTGACCA TGAATTTCTT
4401 TCTTTTATAA ATTTTCTCAT TTAAAAATCA AAAATCTTGC AAAACAAAAA
4451 CCATGTTTCT TTTTCTTGTA TAACTTTTTG TTTTCAGCAA CATAAATTGA
4501 TTTTTAGCTG GCAGACAAGA ATATCCATAT AAGATTTGTT AACCATTTCA
4551 GAGAGTTTGG CAATTTTTAA AAGATAATAA GGTATCATTT TTAAGTATGA
4601 AAATTAACAA TATCCCTGTT GCGCACACTA ATTTTGCATG AGTAAGTTTA
4651 CAAATATGTA TCGTCTGTAA AGCAGCATGT GCAGATTATT CATAATATAG
4701 AAGTTAAAAT AAGTATTAGT GCAATTTTCA GATATTTATT TTTGCACAGA
4751 AAACACATTA TCTGGAGAGA AAGAAAGGAG AATTTTTGAG ACTTGGGTTT
4801 TCTTAATGCC AGTGTGAATT TGCAGATGTT TTCAGAAAAT CAAGTCACAG
4851 TAACAATTTG CCACTTTTTT CTATTATAAA TCTTCTTACT TAAATTTTGA
4901 ATATTTAGTT TTTCTCAGTT ACCCATTTGT GTGTGTGTGA TTCCACTTAG
4951 AAATTCTTAA AACCAGATTT TTCTTTCATT CCGTTTGGAT GTCTACATTC
5001 CTTATCAAAG GATATAAATA CTGTGTATGC TTTTGAATTT TATTTTTAGG
5051 AAAATTCTGA AGCCAGCTAT CACAGGTTTG TTAGCTAATA ATAGTATTTT
5101 CTTTTAGTTG AGTTAGGTTT TTCCCCATCT CCTGTAGAGC GAATTTACAT
5151 ATTGTATTGG GTAAGTGTTC ACTACTTTTC CTGATTAAGG GATCTGTGCT
5201 GGGGAACAAA GCTTTTGCAG TACCTTATAT TGTAGTTAAA ATTTTATTTA
5251 ACATATCCTT CAGTGAGCTC ATTTCACACT GTAGCCTCTT CCTTAAAATT
5301 TGTGGTGCTC CTGTAACAGT AAGAACTAAT TCTGAAATAA AAGACATCTC
5351 CTAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA
```

BLAST Results

Entry HS793345 from database EMBL:
human STS WI-12457.
Score = 1985, P = 1.3e-83, identities = 433/460

Medline entries

97127450:
Molecular cloning of a novel 120-kDa TBP-interacting protein.

Peptide information for frame 2

ORF from 350 bp to 4039 bp; peptide length: 1230

```
Category: known protein
Classification: Nucleic acid management


   1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV
  51 KMILKLLEDK NGEVQNLAVK CLGPLVSKVK EYQVETIVDT LCTNMLSDKE
 101 QLRDISSIGL KTVIGELPPA SSGSALAANV CKKITGRLTS AIAKQEDVSV
 151 QLEALDIMAD MLSRQGGLLV NFHPSILTCL LPQLTSPRLA VRKRTIIALG
 201 HLVMSCGNIV FVDLIEHLLS ELSKNDSMST TRTYIQCIAA ISRQAGHRIG
 251 EYLEKIIPLV VKFCNVDDDE LREYCIQAFE SFVRRCPKEV YPHVSTIINI
 301 CLKYLTYDPN YNYDDEDEDE NAMDADGGDD DDQGSDDEYS DDDDMSWKVR
 351 RAAAKCLDAV VSTRHEMLPE FYKTVSPALI SRFKEREENV KADVFHAYLS
 401 LLKQTRPVQS WLCDPDAMEQ GETPLTMLQS QVPNIVKALH KQMKEKSVKT
 451 RQCCFNMLTE LVNVLPGALT QHIPVLVPGI IFSLNDKSSS SNLKIDALSC
 501 LYVILCNHSP QVFHPHVQAL VPPVVACVGD PFYKITSEAL LVTQQLVKVI
 551 RPLDQPSSFD ATPYIKDLFT CTIKRLKAAD IDQEVKERAI SCMGQIICNL
 601 GDNLGSDLPN TLQIFLERLK NEITRLTTVK ALTLIAGSPL KIDLRPVLGE
 651 GVPILASFLR KNQRALKLGT LSALDILIKN YSDSLTAAMI DAVLDELPPL
 701 ISESDMHVSQ MAISFLTTLA KVYPSSLSKI SGSILNELIG LVRSPLLQGG
 751 ALSAMLDFFQ ALVVTGTNNL GYMDLLRMLT GPVYSQSTAL THKQSYYSIA
 801 KCVAALTRAC PKEGPAVVGQ FIQDVKNSRS TDSIRLLALL SLGEVGHHID
 851 LSGQLELKSV ILEAFSSPSE EVKSAASYAL GSISVGNLPE YLPFVLQEIT
 901 SQPKRQYLLL HSLKEIISSA SVVGLKPYVE NIWALLLKHC ECAEEGTRNV
 951 VAECLGKLTL IDPETLLPRL KGYLISGSSY ARSSVVTAVK FTISDHPQPI
1001 DPLLKNCIGD FLKTLEDPDL NVRRVALVTF NSAAHNKPSL IRDLLDTVLP
1051 HLYNETKVRK ELIREVEMGP FKHTVDDGLD IRKAAFECMY TLLDSCLDRL
1101 DIFEFLNHVE DGLKDHYDIK MLTFLMLVRL STLCPSAVLQ RLDRLVEPLR
1151 ATCTTKVKAN SVKQEFEKQD ELKRSAMRAV AALLTIPEAE KSPLMSEFQS
1201 QISSNPELAA IFESIQKDSS STNLESMDTS

                              BLASTP hits

No BLASTP hits available

          Alert BLASTP hits for DKFZphtes3_22g2, frame 2

TREMBL:AB020636_1 gene: "KIAA0829"; product: "KIAA0829 protein";  Homo
sapiens mRNA for KIAA0829 protein, partial cds., N = 1, Score = 5986, P
= 0

TREMBL:RND6711_1 gene: "tip120"; product: "TIP120";  Rattus norvegicus
mRNA for TIP120, complete cds., N = 1, Score = 6203, P = 0


>TREMBL:RND6711_1 gene: "tip120"; product: "TIP120";  Rattus norvegicus mRNA
     for TIP120, complete cds.
            Length = 1,230

  HSPs:

 Score = 6203 (930.7 bits), Expect = 0.0e+00, P = 0.0e+00
 Identities = 1227/1230 (99%), Positives = 1228/1230 (99%)

Query:      1 MASASYHISNLLEKMTSSDKDFRFMATNDLMTELQKDSIKLDDDSERKVVKMILKLLEDK 60
                MASASYHISNLLEKMTSSDKDFRFMATNDLMTELQKDSIKLDDDSERKVVKMILKLLEDK
Sbjct:      1 MASASYHISNLLEKMTSSDKDFRFMATNDLMTELQKDSIKLDDDSERKVVKMILKLLEDK 60

Query:     61 NGEVQNLAVKCLGPLVSKVKEYQVETIVDTLCTNMLSDKEQLRDISSIGLKTVIGELPPA 120
                NGEVQNLAVKCLGPLVSKVKEYQVETIVDTLCTNMLSDKEQLRDISSIGLKTVIGELPPA
Sbjct:     61 NGEVQNLAVKCLGPLVSKVKEYQVETIVDTLCTNMLSDKEQLRDISSIGLKTVIGELPPA 120

Query:    121 SSGSALAANVCKKITGRLTSAIAKQEDVSVQLEALDIMADMLSRQGGLLVNFHPSILTCL 180
                SSGSALAANVCKKITGRLTSAIAKQEDVSVQLEALDIMADMLSRQGGLLVNFHPSILTCL
Sbjct:    121 SSGSALAANVCKKITGRLTSAIAKQEDVSVQLEALDIMADMLSRQGGLLVNFHPSILTCL 180

Query:    181 LPQLTSPRLAVRKRTIIALGHLVMSCGNIVFVDLIEHLLSELSKNDSMSTTRTYIQCIAA 240
                LPQLTSPRLAVRKRTIIALGHLVMSCGNIVFVDLIEHLLSELSKNDSMSTTRTYIQCIAA
Sbjct:    181 LPQLTSPRLAVRKRTIIALGHLVMSCGNIVFVDLIEHLLSELSKNDSMSTTRTYIQCIAA 240

Query:    241 ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEVYPHVSTIINI 300
                ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEVYPHVSTIINI
Sbjct:    241 ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEVYPHVSTIINI 300

Query:    301 CLKYLTYDPNYNYDDEDEDENAMDADGGDDDDQGSDDEYSDDDDMSWKVRRAAAKCLDAV 360
                CLKYLTYDPNYNYDDEDEDENAMDADGGDDDDQGSDDEYSDDDDMSWKVRRAAAKCLDAV
Sbjct:    301 CLKYLTYDPNYNYDDEDEDENAMDADGGDDDDQGSDDEYSDDDDMSWKVRRAAAKCLDAV 360

Query:    361 VSTRHEMLPEFYKTVSPALISRFKEREENVKADVFHAYLSLLKQTRPVQSWLCDPDAMEQ 420
                VSTRHEMLPEFYKTVSPALISRFKEREENVKADVFHAYLSLLKQTRPVQSWLCDPDAMEQ
Sbjct:    361 VSTRHEMLPEFYKTVSPALISRFKEREENVKADVFHAYLSLLKQTRPVQSWLCDPDAMEQ 420
```

```
Query:   421 GETPLTMLQSQVPNIVKALHKQMKEKSVKTRQCCFNMLTELVNVLPGALTQHIPVLVPGI 480
             GETPLTMLQSQVPNIVKALHKQMKEKSVKTRQCCFNMLTELVNVLPGALTQHIPVLVPGI
Sbjct:   421 GETPLTMLQSQVPNIVKALHKQMKEKSVKTRQCCFNMLTELVNVLPGALTQHIPVLVPGI 480

Query:   481 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPHVQALVPPVVACVGDPFYKITSEAL 540
             IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPHVQALVPPVVACVGDPFYKITSEAL
Sbjct:   481 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPHVQALVPPVVACVGDPFYKITSEAL 540

Query:   541 LVTQQLVKVIRPLDQPSSFDATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL 600
             LVTQQLVKVIRPLDQPSSFDATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL
Sbjct:   541 LVTQQLVKVIRPLDQPSSFDATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL 600

Query:   601 GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPILASFLR 660
             GDNLG DL NTLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPILASFLR
Sbjct:   601 GDNLGPDLSNTLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPILASFLR 660

Query:   661 KNQRALKLGTLSALDILIKNYSDSLTAAMIDAVLDELPPLISESDMHVSQMAISFLTTLA 720
             KNQRALKLGTLSALDILIKNYSDSLTAAMIDAVLDELPPLISESDMHVSQMAISFLTTLA
Sbjct:   661 KNQRALKLGTLSALDILIKNYSDSLTAAMIDAVLDELPPLISESDMHVSQMAISFLTTLA 720

Query:   721 KVYPSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGYMDLLRMLT 780
             KVYPSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGYMDLLRMLT
Sbjct:   721 KVYPSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGYMDLLRMLT 780

Query:   781 GPVYSQSTALTHKQSYYSIAKCVAALTRACPKEGPAVVGQFIQDVKNSRSTDSIRLLALL 840
             GPVYSQSTALTHKQSYYSIAKCVAALTRACPKEGPAVVGQFIQDVKNSRSTDSIRLLALL
Sbjct:   781 GPVYSQSTALTHKQSYYSIAKCVAALTRACPKEGPAVVGQFIQDVKNSRSTDSIRLLALL 840

Query:   841 SLGEVGHHIDLSGQLELKSVILEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT 900
             SLGEVGHHIDLSGQLELKSVILEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT
Sbjct:   841 SLGEVGHHIDLSGQLELKSVILEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT 900

Query:   901 SQPKRQYLLLHSLKEIISSASVVGLKPYVENIWALLLKHCECAEEGTRNVVAECLGKLTL 960
             SQPKRQYLLLHSLKEIISSASVVGLKPYVENIWALLLKHCECAEEGTRNVVAECLGKLTL
Sbjct:   901 SQPKRQYLLLHSLKEIISSASVVGLKPYVENIWALLLKHCECAEEGTRNVVAECLGKLTL 960

Query:   961 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLLKNCIGDFLKTLEDPDL 1020
             IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLLKNCIGDFLKTLEDPDL
Sbjct:   961 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLLKNCIGDFLKTLEDPDL 1020

Query:  1021 NVRRVALVTFNSAAHNKPSLIRDLLDTVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD 1080
             NVRRVALVTFNSAAHNKPSLIRDLLD+VLPHLYNETKVRKELIREVEMGPFKHTVDDGLD
Sbjct:  1021 NVRRVALVTFNSAAHNKPSLIRDLLDSVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD 1080

Query:  1081 IRKAAFECMYTLLDSCLDRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ 1140
             IRKAAFECMYTLLDSCLDRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ
Sbjct:  1081 IRKAAFECMYTLLDSCLDRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ 1140

Query:  1141 RLDRLVEPLRATCTTKVKANSVKQEFEKQDELKRSAMRAVAALLTIPEAEKSPLMSEFQS 1200
             RLDRLVEPLRATCTTKVKANSVKQEFEKQDELKRSAMRAVAALLTIPEAEKSPLMSEFQS
Sbjct:  1141 RLDRLVEPLRATCTTKVKANSVKQEFEKQDELKRSAMRAVAALLTIPEAEKSPLMSEFQS 1200

Query:  1201 QISSNPELAAIFESIQKDSSSTNLESMDTS 1230
             QISSNPELAAIFESIQKDSSSTNLESMDTS
Sbjct:  1201 QISSNPELAAIFESIQKDSSSTNLESMDTS 1230
```

Pedant information for DKFZphtes3_22g2, frame 2

Report for DKFZphtes3_22g2.2

```
[LENGTH]        1230
[MW]            136376.58
[pI]            5.52
[HOMOL]         TREMBL:RND6711_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for
TIP120, complete cds. 0.0
[KW]            TRANSMEMBRANE  1
[KW]            LOW_COMPLEXITY      5.28 %

SEQ     MASASYHISNLLEKMTSSDKDFRFMATNDLMTELQKDSIKLDDDSERKVVKMILKLLEDK
SEG     ............................................................
PRD     cccccchhhhhhhhccccceeeeehhhhhhhhhccccccccchhhhhhhhhhhhcc
MEM     ............................................................

SEQ     NGEVQNLAVKCLGPLVSKVKEYQVETIVDTLCTNMLSDKEQLRDISSIGLKTVIGELPPA
SEG     ........................................................xxxx
PRD     cccccceeeeeeeceeeeehhhhhhhhhhhhcccchhhhhcccccccchhhhhhhhhcccc
```

```
MEM     ............................................................

SEQ     SSGSALAANVCKKITGRLTSAIAKQEDVSVQLEALDIMADMLSRQGGLLVNFHPSILTCL
SEG     xxxxxxxx....................................................
PRD     cccccchhhhhhcchhhhhhhccccchhhhhhhhhhhhhhhhhhhcccceeeecchhhhhh
MEM     ............................................................

SEQ     LPQLTSPRLAVRKRTIIALGHLVMSCGNIVFVDLIEHLLSELSKNDSMSTTRTYIQCIAA
SEG     ............................................................
PRD     hcccccchhhhhhhhhhhheeeeeccccceeehhhhhhhhhhhcccccchhhhhhhhhhh
MEM     ...............MMMMMMMMMMMMMMMMM............................

SEQ     ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEVYPHVSTIINI
SEG     ............................................................
PRD     hhhhcccccccchhhhhhhhheeeeeccchhhhhhhhhhhhhhhhhcccccceeecchhhhhh
MEM     ............................................................

SEQ     CLKYLTYDPNYNYDDEDEDENAMDADGGDDDDQGSDDEYSDDDDMSWKVRRAAAKCLDAV
SEG     .........xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx..............
PRD     hhhhhcccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhh
MEM     ............................................................

SEQ     VSTRHEMLPEFYKTVSPALISRFKEREENVKADVFHAYLSLLKQTRPVQSWLCDPDAMEQ
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeccccccc
MEM     ............................................................

SEQ     GETPLTMLQSQVPNIVKALHKQMKEKSVKTRQCCFNMLTELVNVLPGALTQHIPVLVPGI
SEG     ............................................................
PRD     ccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhcccccccccceeeecce
MEM     ............................................................

SEQ     IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPHVQALVPPVVACVGDPFYKITSEAL
SEG     xxxxxxxxxxxxxxx.............................................
PRD     eeeecccccccccchhhhhhhheeeeecccccccccceeeeeecceeeeeeccchhhhhhhh
MEM     ............................................................

SEQ     LVTQQLVKVIRPLDQPSSFDATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL
SEG     ............................................................
PRD     hhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhcchhhhhhhhhhhhhheeeecc
MEM     ............................................................

SEQ     GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPILASFLR
SEG     ............................................................
PRD     cccccccccchhhhhhhhhcchhhhhhhhhhhhheeeecccccccccceeehhhhhhhhhh
MEM     ............................................................

SEQ     KNQRALKLGTLSALDILIKNYSDSLTAAMIDAVLDELPPLISESDMHVSQMAISFLTTLA
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhcccccccchhhhhhhhhhhhhhhh
MEM     ............................................................

SEQ     KVYPSSLSKISGSILNELIGLVRSPLLQGGALSAMLDFFQALVVTGTNNLGYMDLLRMLT
SEG     ............................................................
PRD     cccccceeecchhhhhhhhhhhcccccccchhhhhhhhhhhheeeeccccccchhhhhhhhc
MEM     ............................................................

SEQ     GPVYSQSTALTHKQSYYSIAKCVAALTRACPKEGPAVVGQFIQDVKNSRSTDSIRLLALL
SEG     ............................................................
PRD     ccccccccchhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhcccccchhhhhhhh
MEM     ............................................................

SEQ     SLGEVGHHIDLSGQLELKSVILEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT
SEG     ............................................................
PRD     hccccccccccccccccceeeeeeccccchhhhhhhhhhhcccccccccccchhhhhhhhh
MEM     ............................................................

SEQ     SQPKRQYLLLHSLKEIISSASVVGLKPYVENIWALLLKHCECAEEGTRNVVAECLGKLTL
SEG     ............................................................
PRD     ccchhhhhhhhhhhhhhhccccceeehhhhhhhhhhhhhhhhhccccccceeeeeecccccccc
MEM     ............................................................

SEQ     IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLLKNCIGDFLKTLEDPDL
SEG     ............................................................
PRD     cccccccccccccccccccccchhhhhhhhhhcccccccccccchhhhhhhhhhhhccccc
MEM     ............................................................

SEQ     NVRRVALVTFNSAAHNKPSLIRDLLDTVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD
SEG     ............................................................
PRD     ccceeeeeeecccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccch
MEM     ............................................................
```

```
SEQ    IRKAAFECMYTLLDSCLDRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhcccccccceeeeccccccccccchhhhhhhhhhhhhhhhhccccchhhh
MEM    ............................................................

SEQ    RLDRLVEPLRATCTTKVKANSVKQEFEKQDELKRSAMRAVAALLTIPEAEKSPLMSEFQS
SEG    ............................................................
PRD    hhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccchhhhh
MEM    ............................................................

SEQ    QISSNPELAAIFESIQKDSSSTNLESMDTS
SEG    ..............................
PRD    hhhcccchhhhhhhhhhhhccccccccccccc
MEM    ..............................
```

(No Prosite data available for DKFZphtes3_22g2.2)

(No Pfam data available for DKFZphtes3_22g2.2)

DKFZphtes3_22n13

group: testes derived

DKFZphtes3_22n13 encodes a novel 677 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

dJ1042K10.3, complete

Sequenced by LMU

Locus: /map="22q13.1-13.2"

Insert length: 3353 bp
Poly A stretch at pos. 3315, polyadenylation signal at pos. 3298

```
   1 ATGGAACCAC TATCCCCACT GCCAAGTCCA CCCCCACACT CATTAAGCAA
  51 AGCCAACCCA AGTCTGCCAG TGAGAAGTCA CAGCGCAGCA AGAAGGCCAA
 101 GGAGCTGAAG CCAAAGGTGA AGAAGCTCAA GTACCACCAG TACATCCCCC
 151 CGGACCAGAA GCAGGACAGG GGGGCACCCC CCATGGACTC ATCCTACGCC
 201 AAGATCCTGC AGCAGCAGCA GCTCTTCCTC CAGCTGCAGA TCCTCAACCA
 251 GCAGCAGCAG CAGCACCACA ACTACCAGGC CATCCTGCCT GCCCCGCCAA
 301 AGTCAGCAGG CGAGGCCCTG GGAAGCAGCG GGACCCCCCC AGTACGCAGC
 351 CTCTCCACTA CCAATAGCAG CTCCAGCTCG GGCGCCCCTG GGCCCTGTGG
 401 GCTGGCACGT CAGAACAGCA CCTCACTGAC TGGCAAGCCG GGAGCCCTGC
 451 CGGCCAACCT GGACGACATG AAGGTGGCAG AGCTGAAGCA GGAGCTGAAG
 501 TTGCGATCAC TGCCTGTCTC GGGCACCAAA ACTGAGCTGA TTGAGCGCCT
 551 TCGAGCCTAT CAAGACCAAA TCAGCCCTGT GCCAGGAGCC CCCAAGGCCC
 601 CTGCCGCCAC CTCTATCCTG CACAAGGCTG GCGAGGTGGT GGTAGCCTTC
 651 CCAGCGGCCC GGCTGAGCAC GGGGCCAGCC CTGGTGGCAG CAGGCCTGGC
 701 TCCAGCTGAG GTGGTGGTGG CCACGGTGGC CAGCAGTGGG GTGGTGAAGT
 751 TTGGCAGCAC GGGCTCCACG CCCCCCGTGT CTCCCACCCC CTCGGAGCGC
 801 TCACTGCTCA GCACGGGCGA TGAAAACTCC ACCCCCGGGG ACACCTTTGG
 851 TGAGATGGTG ACATCACCTC TGACGCAGCT GACCCTGCAG GCCTCGCCAC
 901 TGCAGATCCT CGTGAAGGAG GAGGGCCCCC GGGCCGGGTC CTGTTGCCTG
 951 AGCCCTGGGG GGCGGGCGGA GCTAGAGGGG CGCGACAAGG ACCAGATGCT
1001 GCAGGAGAAA GACAAGCAGA TCGAGGCGCT GACGCGCATG CTCCGGCAGA
1051 AGCAGCAGCT GGTGGAGCGG CTCAAGCTGC AGCTGGAGCA GGAGAAGCGA
1101 GCCCAGCAGC CCGCCCCCGC CCCCGCCCCC CTCGGCACCC CCGTGAAGCA
1151 GGAGAACAGC TTCTCCAGCT GCCAGCTGAG CCAGCAGCCC CTGGGCCCCG
1201 CTCACCCATT CAACCCCAGC CTGGCGGCCC CAGCCACCAA CCACATAGAC
1251 CCTTGTGCTG TGGCCCCAGG GCCCCCGTCC GTGGTGGTGA AGCAGGAAGC
1301 CTTGCAGCCT GAGCCCGAGC CGGTCCCCGC CCCCCAGTTG CTTCTGGGGC
1351 CTCAGGGCCC CGGCCTCATC AAGGGGGTTG CACCTCCCAC CCTCATCACC
1401 GACTCCACAG GGACCCACCT TGTCCTCACC GTGACCAATA AGAATGCAGA
1451 CAGCCCTGGC CTGTCCAGTG GGAGCCCCCA GCAGCCCTCG TCCCAGCCTG
1501 GCTCTCCAGC GCCTGCCCCC TCTGCCCAGA TGGACCTGGA GCACCCACTG
1551 CAGCCCCTCT TTGGGACCCC CACTTCTCTG CTGAAGAAGG AACCACCTGG
1601 CTATGAGGAA GCCATGAGCC AGCAGCCCAA ACAGCAGGAA AATGGTTCCT
1651 CAAGCCAGCA GATGGACGAC CTGTTTGACA TTCTCATTCA GAGCGGAGAA
1701 ATTTCAGCAG ATTTCAAGGA GCCGCCATCC CTGCCAGGGA AGGAGAAGCC
1751 ATCCCCGAAG ACAGTCTGTG GGTCCCCCCT GGCAGCACAG CCATCACCTT
1801 CTGCTGAGCT CCCCCAGGCT GCCCCACCTC CTCCAGGCTC ACCCTCCCTC
1851 CCTGGACGCC TGGAGGACTT CCTGGAGAGC AGCACGGGGC TGCCCCTGCT
1901 GACCAGTGGG CATGACGGGC CAGAGCCCCT TTCCCTCATT GACGACCTCC
1951 ATAGCCAGAT GCTGAGCAGC ACTGCCATCC TGGACCACCC CCCGTCACCC
2001 ATGGACACCT CGGAATTGCA CTTTGTTCCT GAGCCCAGCA GCACCATGGG
2051 CCTGGACCTG GCTGATGGCC ACCTGGACAG CATGGACTGG CTGGAGCTGT
2101 CGTCAGGTGG TCCCGTGCTG AGCCTAGCCC CCCTCAGCAC CACAGCCCCC
2151 AGCCTCTTCT CCACAGACTT CCTCGATGGC CATGATTTGC AGCTGCACTG
2201 GGATTCCTGC TTGTAGCTCT CTGGCTCAAG ACGGGGTGGG GAAGGGGCTG
2251 GGAGCCAGGG TACTCCAATG CGTGGCTCTC CTGCGTGATT CGGCCTCTCC
2301 ACATGGTTGT GAGTCTTGAC AATCACAGCC CCTGCTTTTT CCCTTCCCTG
2351 GGAGGCTAGA ACAGAGAAGC CCTTACTCCT GGTTCAGTGC CACGCAGGGC
2401 AGAGGAGAGC AGCTGTCAAG AAGCAGCCCT GGCTCTCACG CTGGGGTTTT
2451 GGACACACGG TCAGGGTCAG GGCCATTTCA GCTTGACCTC CTTTTTTGAG
2501 GTCAGGGGGC ACTGTCTGTC TGGCTACAAT TTGGCTAAGG TAGGTGAAGC
2551 CTGGCCAGGC GGGAGGCTTC TCTTCTGACC CAGGGCTGAG ACAGGTTAAG
2601 GGGTGAATCT CCTTCCTTTC TCTCCCTGCT TTGCTGTGAA GGGAGAAATT
2651 AGCCTGGGCC TCTACCCCCT ATTCCCTGTG TCTGCCAACC CCAGGATCCC
2701 AGGGCTCCCT GCCATTTTAG TGTCTTGGTG TAGTGTAACC ATTTAGTGGT
2751 TGGTGGCAAC AATTTTATGT ACAGGTGTAT ATACCTCTAT ATTATATATC
2801 GACATACATA TATATTTTTG GGGGGGGGCG GACAGGAGAT GGGTGCAACT
```

```
2851 CCCTCCCATC CTACTCTCAC AGAAGGGCCT GGATGCAAGG TTACCCTTGA
2901 GCTGTGTGCC ACAGTCTGGT GCCCAGTCTG GCATGCAGCT ACCCAGGCCC
2951 ACCCATCACG TGTGATTGAC ATGTAGGTAC CCTGCCACGG CCTATGCCCC
3001 ACCTGCCCTG CTTCCTGGCT CCTTATCAGT GCCATGAGGG CAGAGGTGCT
3051 ACCTGGCCTT CCTGCCAGGA GCTCTCCACC CACTCACATT CCGTCCCCGC
3101 CGCCTCACTG CAGCCAGCGT GGCCCTAGGA CAGGAGGAGC TTCGGGCCCA
3151 GCTTCACCCT GCGGTGGGGC TGAGGGGTGG CCATCTCCTG CCCTGGGGCC
3201 ACTGGCTTCA CATTCTGGGC TGACTCATAG GGGAGTAGGG GTGGAGTCAC
3251 CAAAACCAGT GCTGGGACAA AGATGGGGAA GGTGTGTGAA CTTTTTAAAA
3301 TAAACACAAA AACACAGGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAG
```

BLAST Results

Entry HS1042K10 from database EMBL:
Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2.
Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2,
Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP
domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a
putative CpG island.
Score = 7997, P = 0.0e+00, identities = 1617/1645
7 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 183 bp to 2213 bp; peptide length: 677
Category: similarity to unknown protein
Classification: unclassified

```
  1 MDSSYAKILQ QQQLFLQLQI LNQQQQQHHN YQAILPAPPK SAGEALGSSG
 51 TPPVRSLSTT NSSSSSGAPG PCGLARQNST SLTGKPGALP ANLDDMKVAE
101 LKQELKLRSL PVSGTKTELI ERLRAYQDQI SPVPGAPKAP AATSILHKAG
151 EVVVAFPAAR LSTGPALVAA GLAPAEVVVA TVASSGVVKF GSTGSTPPVS
201 PTPSERSLLS TGDENSTPGD TFGEMVTSPL TQLTLQASPL QILVKEEGPR
251 AGSCCLSPGG RAELEGRDKD QMLQEKDKQI EALTRMLRQK QQLVERLKLQ
301 LEQEKRAQQP APAPAPLGTP VKQENSFSSC QLSQQPLGPA HPFNPSLAAP
351 ATNHIDPCAV APGPPSVVVK QEALQPEPEP VPAPQLLLGP QGPGLIKGVA
401 PPTLITDSTG THLVLTVTNK NADSPGLSSG SPQQPSSQPG SPAPAPSAQM
451 DLEHPLQPLF GTPTSLLKKE PPGYEEAMSQ QPKQQENGSS SQQMDDLFDI
501 LIQSGEISAD FKEPPSLPGK EKPSPKTVCG SPLAAQPSPS AELPQAAPPP
551 PGSPSLPGRL EDFLESSTGL PLLTSGHDGP EPLSLIDDLH SQMLSSTAIL
601 DHPPSPMDTS ELHFVPEPSS TMGLDLADGH LDSMDWLELS SGGPVLSLAP
651 LSTTAPSLFS TDFLDGHDLQ LHWDSCL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_22n13, frame 3

TREMBL:HS1042K10_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)"; Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative CpG island., N = 1, Score = 1285, P = 4.9e-131

TREMBL:CEUK06A9_3 gene: "K06A9.1a"; Caenorhabditis elegans cosmid K06A9., N = 2, Score = 149, P = 1.3e-09

TREMBLNEW:SSI132828_1 product: "p210 protein"; Spermatozopsis similis mRNA for p210 protein, partial, N = 1, Score = 171, P = 2.8e-09

>TREMBL:HS1042K10_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel
protein)"; Human DNA sequence from clone 1042K10 on chromosome
22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC

```
     4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP
     domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a
     putative CpG island.
            Length = 243

  HSPs:

 Score = 1285 (192.8 bits), Expect = 4.9e-131, P = 4.9e-131
 Identities = 243/243 (100%), Positives = 243/243 (100%)

Query:   435 PSSQPGSPAPAPSAQMDLEHPLQPLFGTPTSLLKKEPPGYEEAMSQQPKQQENGSSSQQM 494
             PSSQPGSPAPAPSAQMDLEHPLQPLFGTPTSLLKKEPPGYEEAMSQQPKQQENGSSSQQM
Sbjct:     1 PSSQPGSPAPAPSAQMDLEHPLQPLFGTPTSLLKKEPPGYEEAMSQQPKQQENGSSSQQM 60

Query:   495 DDLFDILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSPSAELPQAAPPPPGSP 554
             DDLFDILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSPSAELPQAAPPPPGSP
Sbjct:    61 DDLFDILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSPSAELPQAAPPPPGSP 120

Query:   555 SLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAILDHPPSPMDTSELHF 614
             SLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAILDHPPSPMDTSELHF
Sbjct:   121 SLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAILDHPPSPMDTSELHF 180

Query:   615 VPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDLQLHWD 674
             VPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDLQLHWD
Sbjct:   181 VPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDLQLHWD 240

Query:   675 SCL 677
             SCL
Sbjct:   241 SCL 243
```

Pedant information for DKFZphtes3_22n13, frame 3

Report for DKFZphtes3_22n13.3

```
[LENGTH]       677
[MW]           70743.01
[pI]           4.93
[HOMOL]        TREMBL:HS1042K10_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)";
Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for
Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with
probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative
CpG island. 1e-111
[KW]           TRANSMEMBRANE  1
[KW]           LOW_COMPLEXITY    21.57 %
[KW]           COILED_COIL        4.58 %

SEQ     MDSSYAKILQQQQLFLQLQILNQQQQQHHNYQAILPAPPKSAGEALGSSGTPPVRSLSTT
SEG     ........xxxxxxxxxxxxxxxxxxx............................xxxxx
PRD     ccchhhhhhhhhhhhhhhhhhhhhhhhcceeeeeeccccccceeeeccccccccceeecccc
COILS   ............................................................
MEM     ............................................................

SEQ     NSSSSSGAPGPCGLARQNSTSLTGKPGALPANLDDMKVAELKQELKLRSLPVSGTKTELI
SEG     xxxxxx......................................................
PRD     cccccccccccceeeccccccccccccccccccccchhhhhhhhhhhhhhccccccchhhhh
COILS   ............................................................
MEM     ............................................................

SEQ     ERLRAYQDQISPVPGAPKAPAATSILHKAGEVVVAFPAARLSTGPALVAAGLAPAEVVVA
SEG     ...........................................xxxxxxxxxxxxxxxxx
PRD     hhhhhhhhcccccccccccccceeeeeeecccceeeecccccccccccccccccccceeeeee
COILS   ............................................................
MEM     ......................................MMMMMMMMMMMMMMMMMMMMMM

SEQ     TVASSGVVKFGSTGSTPPVSPTPSERSLLSTGDENSTPGDTFGEMVTSPLTQLTLQASPL
SEG     xxxxxxxx..xxxxxxxxxxxxxx....................................
PRD     eeeccccccccccccccccccccceeeecccccccccccccccccceeeccccceeeecccce
COILS   ............................................................
MEM     M...........................................................

SEQ     QILVKEEGPRAGSCCLSPGGRAELEGRDKDQMLQEKDKQIEALTRMLRQKQQLVERLKLQ
SEG     ............................................................
PRD     eeeeccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ....................................CCCCCCCCCCCCCCCCCCCCCCCC
MEM     ............................................................

SEQ     LEQEKRAQQPAPAPAPLGTPVKQENSFSSCQLSQQPLGPAHPFNPSLAAPATNHIDPCAV
```

```
SEG     ......xxxxxxxxxx............................................
PRD     hhhhhhhhcccccccccccccccccccceeeeeccccccccccccceeeccccccccccccc
COILS   CCCCCCC.....................................................
MEM     ............................................................

SEQ     APGPPSVVVKQEALQPEPEPVPAPQLLLGPQGPGLIKGVAPPTLITDSTGTHLVLTVTNK
SEG     .......................xxxxxxxxxxxx.........................
PRD     cccccceeeeeccccccccccccceeecccccccceeeeecccccccccceeeeeeeeecc
COILS   ............................................................
MEM     ............................................................

SEQ     NADSPGLSSGSPQQPSSQPGSPAPAPSAQMDLEHPLQPLFGTPTSLLKKEPPGYEEAMSQ
SEG     ...xxxxxxxxxxxxxxxxxxxxxxxxxx...............................
PRD     ccccccccccccccccccccccccccchhhhhhhhhcccccccccccccccccccccccc
COILS   ............................................................
MEM     ............................................................

SEQ     QPKQQENGSSSQQMDDLFDILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSPS
SEG     .................................................xxxxxxxxxxx
PRD     cccccccccccchhhhhhhhhhcccccccccccccccccccccccccccccccccccccc
COILS   ............................................................
MEM     ............................................................

SEQ     AELPQAAPPPPGSPSLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAIL
SEG     xxxxxxxxxxxxxxxxxx..........................................
PRD     cccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhcccceee
COILS   ............................................................
MEM     ............................................................

SEQ     DHPPSPMDTSELHFVPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFS
SEG     ............................................................
PRD     ccccccccccccccccccccccccccccccccccceeeeccccceeeeeecccccccccc
COILS   ............................................................
MEM     ............................................................

SEQ     TDFLDGHDLQLHWDSCL
SEG     .................
PRD     cccccccceeecccccc
COILS   .................
MEM     .................
```

(No Prosite data available for DKFZphtes3_22n13.3)

(No Pfam data available for DKFZphtes3_22n13.3)

DKFZphtes3_23111
----

group: intracellular transport and trafficking

DKFZphtes3_23111 encodes a novel 186 amino acid protein nearly identical to mouse ADP-ribosylation-like factor homolog 6 (Arl6).

Protein secretion through the endoplasmic reticulum and the Golgi vesicular trafficking system is initiated by the binding of ADP-ribosylation factors (ARFs) to donor membranes, leading to recruitment of cocatomer, bud formation, and eventual vesicle release. ARFs are approximately 20-kDa GTPases that are active with bound GTP and inactive with GDP bound. The novel protein contains an ATP/GTP-binding site motif A (P-loop) and seems to be a novel ARF. It seems to have an important role in vesicular transport and vesicular trafficking.

The new protein can find application in modulating vesicle transport and trafficking in cells.

nearly identical to mouse Arl6, ADP-ribosylation-like factor homolog

start at Bp 15 matches kozak consensus ANNatgG

Sequenced by LMU

Locus: unknown

Insert length: 717 bp
Poly A stretch at pos. 689, no polyadenylation signal found

```
  1 ATTTGAATCA CATTATGGGA TTGCTAGACA GACTTTCAGT CTTGCTTGGC
 51 CTGAAGAAGA AGGAGGTTCA TGTTTTGTGC CTTGGGCTAG ATAATAGTGG
101 CAAAACGACG ATCATTAACA AACTTAAACC TTCAAATGCT CAATCTCAAA
151 ATATCCTTCC AACAATAGGA TTCAGCATAG AGAAATTCAA ATCATCCAGT
201 TTGTCATTTA CAGTGTTTGA CATGTCAGGT CAAGGAAGAT ACAGAAATCT
251 CTGGGAACAC TATTATAAAG AAGGCCAAGC TATTATTTTT GTCATTGATA
301 GTAGTGATAG ATTAAGAATG GTTGTGGCCA AAGAAGAACT CGATACTCTT
351 CTGAATCATC CAGATATTAA ACACCGTCGA ATTCCAATCT TATTCTTTGC
401 AAATAAAATG GATCTTAGAG ATGCAGTGAC ATCTGTAAAA GTGTCTCAGT
451 TGCTGTGTTT AGAGAACATC AAAGATAAAC CCTGGCATAT TTGTGCTAGT
501 GATGCCATAA AAGGAGAAGG CTTGCAAGAA GGTGTAGACT GGCTTCAAGA
551 TCAGATCCAG ACTGTGAAGA CATGAAAAGA TAATAGTTGG AAACCTCAGC
601 AATTTTCAAT TCAAGGAATC TATCTAAGAC AAATAGAATA CATTTTGTAA
651 AAGATGTTTA TGCATCAAAA AATATAATTT TCTGCTTGCA AAAAAAAAAA
701 AAAAAAAAAA AAAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 15 bp to 572 bp; peptide length: 186
Category: strong similarity to known protein
Classification: Intacellular transport and traffic
Prosite motifs: ATP_GTP_A (24-32)

```
  1 MGLLDRLSVL LGLKKKEVHV LCLGLDNSGK TTIINKLKPS NAQSQNILPT
 51 IGFSIEKFKS SSLSFTVFDM SGQGRYRNLW EHYYKEGQAI IFVIDSSDRL
101 RMVVAKEELD TLLNHPDIKH RRIPILFFAN KMDLRDAVTS VKVSQLLCLE
151 NIKDKPWHIC ASDAIKGEGL QEGVDWLQDQ IQTVKT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_23111, frame 3

TREMBL:AF031903_1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds., N = 1, Score = 923, P = 1.1e-92

TREMBL:CEC38D4_5 gene: "C38D4.8"; Caenorhabditis elegans cosmid C38D4, N = 1, Score = 418, P = 3.6e-39

PIR:S66337 ADP-ribosylation factor 1 - Chlamydomonas reinhardtii, N = 1, Score = 373, P = 2.1e-34

SWISSPROT:ARF1_CHLRE ADP-RIBOSYLATION FACTOR 1., N = 1, Score = 372, P = 2.7e-34

```
>TREMBL:AF031903_1 gene: "Arl6"; product: "ADP-ribosylation-like factor
     homolog ARL6";  Mus musculus ADP-ribosylation-like factor homolog ARL6
     (Arl6) mRNA, complete cds.
            Length = 186

  HSPs:

 Score = 923 (138.5 bits), Expect = 1.1e-92, P = 1.1e-92
 Identities = 178/186 (95%), Positives = 184/186 (98%)

Query:      1 MGLLDRLSVLLGLKKKEVHVLCLGLDNSGKTTIINKLKPSNAQSQNILPTIGFSIEKFKS 60
              MGLLDRLS LLGLKKKEVHVLCLGLDNSGKTTIINKLKPSNAQSQ+I+PTIGFSIEKFKS
Sbjct:      1 MGLLDRLSGLLGLKKKEVHVLCLGLDNSGKTTIINKLKPSNAQSQDIVPTIGFSIEKFKS 60

Query:     61 SSLSFTVFDMSGQGRYRNLWEHYYKEGQAIIFVIDSSDRLRMVVAKEELDTLLNHPDIKH 120
              SSLSFTVFDMSGQGRYRNLWEHYYK+GQAIIFVIDSSD+LRMVVAKEELDTLLNHPDIKH
Sbjct:     61 SSLSFTVFDMSGQGRYRNLWEHYYKDGQAIIFVIDSSDKLRMVVAKEELDTLLNHPDIKH 120

Query:    121 RRIPILFFANKMDLRDAVTSVKVSQLLCLENIKDKPWHICASDAIKGEGLQEGVDWLQDQ 180
              RRIPILFFANKMDLRD+VTSVKVSQLLCLE+IKDKPWHICASDAIKGEGLQEGVDWLQDQ
Sbjct:    121 RRIPILFFANKMDLRDSVTSVKVSQLLCLESIKDKPWHICASDAIKGEGLQEGVDWLQDQ 180

Query:    181 IQTVKT 186
              IQ VKT
Sbjct:    181 IQAVKT 186
```

Pedant information for DKFZphtes3_23111, frame 3

Report for DKFZphtes3_23111.3

[LENGTH] 186
[MW] 21097.69
[pI] 8.72
[HOMOL] TREMBL:AF031903_1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds. 4e-94
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YDL192w] 1e-36
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDL192w] 1e-36
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL192w] 1e-36
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL137w] 2e-36
[FUNCAT] 06.07 protein modification (glycolsylation, acylation, myristylation, palmitylation, farnesylation and processing) [S. cerevisiae, YBR164c] 2e-32
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBR164c] 2e-32
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YMR138w] 4e-19
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YMR138w] 4e-19
[FUNCAT] r general function prediction [M. jannaschii, MJ1339] 2e-05
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YHR005c] 4e-05
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YHR005c] 4e-05
[FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YHR005c] 4e-05
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YKR014c] 2e-04
[FUNCAT] 08.19 cellular import [S. cerevisiae, YKR014c] 2e-04
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YKR014c] 2e-04
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w] 4e-04
[BLOCKS] BL01288C
[BLOCKS] BL01020C SAR1 family proteins
[BLOCKS] BL01019C ADP-ribosylation factors family proteins

```
[BLOCKS]      BL01019B ADP-ribosylation factors family proteins
[BLOCKS]      BL01019A ADP-ribosylation factors family proteins
[SCOP]        d1as3_2 3.29.1.4.12 Transducin (alpha subunit), insertion domai 2e-45
[SCOP]        d1mh1__ 3.29.1.4.2 Rac1  [Human (Homo sapiens)  2e-46
[SCOP]        d5p21__ 3.29.1.4.1 cH-p21 Ras protein  [human (Homo sapiens)  5e-37
[SCOP]        d1hura_ 3.29.1.4.8 ADP-ribosylation factor 1 (ARF1)  [human (Hom 4e-61
[SCOP]        d1a2kc_ 3.29.1.4.5 Ran Nuclear transport factor-2 (NTF2)  [Do 4e-33
[PIRKW]       glycoprotein 2e-33
[PIRKW]       monomer 3e-31
[PIRKW]       P-loop 2e-35
[PIRKW]       lipoprotein 2e-33
[PIRKW]       GTP binding 2e-35
[SUPFAM]      ADP-ribosylation factor 2e-35
[PROSITE]     ATP_GTP_A      1
[PFAM]        ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
[KW]          Alpha_Beta
[KW]          3D
[KW]          LOW_COMPLEXITY      5.91 %

SEQ    MGLLDRLSVLLGLKKKEVHVLCLGLDNSGKTTIINKLKPSNAQSQNILPTIGFSIEKFKS
SEG    ..xxxxxxxxxxx...............................................
1hurA  ..............CCCCEEEEEETTTTCHHHHHHHHCCCCEEEE--EEETTEEEEEEEE

SEQ    SSLSFTVFDMSGQGRYRNLWEHYYKEGQAIIFVIDSSDRLRMVVAKEELDTLLNHPDIKH
SEG    ............................................................
1hurA  TTEEEEEEETTTTTTTCCCHHHHHHCEEEEEEEEETTTTTHHHHHHHHHHHHHHTTTT--

SEQ    RRIPILFFANKMDLRDAVTSVKVSQLLCLENIKDKPWHICASDAIKGEGLQEGVDWLQDQ
SEG    ............................................................
1hurA  TTTEEEEEEETTTTTTTCCHHHHHHHHCGGGTTTTCEEEEECBTTTTBTHHHHHHHHHHH

SEQ    IQTVKT
SEG    ......
1hurA  HHHHC.
```

Prosite for DKFZphtes3_23111.3

```
PS00017        24->32   ATP_GTP_A           PDOC00017
```

Pfam for DKFZphtes3_23111.3

```
HMM_NAME      ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)

HMM              *GMgWfsIFrkMWGlWNKEMRILMLGLDNAGKTTILYMLKlgE..IVTTI
                   MG++  ++ ++GL +KE+++L LGLDN+GKTTI+++LK+       ++
Query          1  -MGLLDRLSVLLGLKKKEVHVLCLGLDNSGKTTIINKLKPSNAQSQNIL    48

HMM               PTIGFNVETVeYKNIKFNVWDVGGQdsIRPYWRHYYpNTDGIIWVVDSaD
                  PTIGF +E+ +    ++F+V+D  GQ + R +W HYY + ++II+V+DS+D
Query         49  PTIGFSIEKFKSSSLSFTVFDMSGQGRYRNLWEHYYKEGQAIIFVIDSSD    98

HMM               RDRMeEaKqELHaMLNEEEL..rDAPlLIFANKQDLPgAMSesEIREaLG
                  R RM  AK+EL+ +LN+ ++  R+ P+L FANK DL++A+++ +++ +L
Query         99  RLRMVVAKEELDTLLNHPDIKHRRIPILFFANKMDLRDAVTSVKVSQLLC   148

HMM               LHeIRCnRPWYIQMCCAVtGEGLYEGMDWLSNYInkRkK*
                  L++I+ + PW+I +++A++GEGL+EG DWL ++I+  K
Query        149  LENIK-DKPWHICASDAIKGEGLQEGVDWLQDQIQTVKT    186
```

DKFZphtes3_23n19
----

group: testes derived

DKFZphtes3_23n19 encodes a novel 387 amino acid protein with similarity to rat protein kinase C-interacting RBCC protein 1.

The novel protein contains not the RING-B box-coiled coil (RBCC) motif of RBCC protein 1, and thus is not a member of this subgroup of RING finger proteins.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to rat protein kinase C-interacting RBCC protein 1

start at Bp 209 matches kozak consensus PyNNatgG
similarity to of C-terminal part to N-terminus of RBCK1

Sequenced by LMU

Locus: unknown

Insert length: 1579 bp
Poly A stretch at pos. 1535, polyadenylation signal at pos. 1515

```
   1 CGGAGACCCT CGGGCCGTGT CCATTTGTGG GCAAAGCCAG CGGGGCAGGC
  51 TTGGCCAGAG TGCACCACTC GGCGCCGTCC CAGGCCCGAC GCTCTGGGCG
 101 CGCCCGGAAC CCCAGGTTCG CGGCCCGTGT TTCCGACCGG CGGAGGGGGC
 151 TCAGCGGCCC GATCCCACGG AAGCGCGCTC GGAGGGGTGG GACCCGGCCG
 201 GACCGGAGAT GGCGCCGCCA GCGGGCGGGG CGGCGGCGGC GGCCTCGGAC
 251 TTGGGCTCCG CCGCAGTGCT CTTGGCTGTG CACGCCGCGG TGAGGCCGCT
 301 GGGCGCCGGG CCAGACGCCG AGGCACAGCT GCGGAGGCTG CAGCTGAGCG
 351 CGGACCCTGA GAGGCCTGGG CGCTTCCGGC TGGAGCTGCT GGGCGCGGGA
 401 CCTGGGGCGG TTAATTTGGA GTGGCCCCTG GAGTCAGTTT CCTACACCAT
 451 CCGAGGCCCC ACCCAGCACG AGCTACAGCC TCCACCAGGA GGGCCTGGAA
 501 CCCTCAGCCT GCACTTCCTC AACCCTCAGG AAGCTCAGCG GTGGGCAGTC
 551 CTAGTCCGAG GTGCCACCGT GGAAGGACAG AATGGCAGCA AGAGCAACTC
 601 ACCACCAGCC TTGGGCCCAG AAGCATGCCC TGTCTCCCTG CCCAGTCCCC
 651 CGGAAGCCTC CACACTCAAG GGCCCTCCAC CTGAGGCAGA TCTTCCTAGG
 701 AGCCCTGGAA ACTTGACGGA GAGAGAAGAG CTGGCAGGGA GCCTGGCCCG
 751 GGCTATTGCA GGTGGAGACG AGAAGGGGGC AGCCCAAGTG GCAGCCGTCC
 801 TGGCCCAGCA TCGTGTGGCC CTGAGTGTTC AGCTTCAGGA GGCCTGCTTC
 851 CCACCTGGCC CCATCAGGCT GCAGGTCACA CTTGAAGACG CTGCCTCTGC
 901 CGCATCCGCC GCGTCCTCTG CACACGTTGC CCTGCAGGTC CACCCCCACT
 951 GCACTGTTGC AGCTCTCCAG GAGCAGGTGT TCTCAGAGCT CGGTTTCCCG
1001 CCAGCCGTGC AACGCTGGGT CATCGGACGG TGCCTGTGTG TGCCTGAGCG
1051 CAGCCTTGCC TCTTACGGGG TTCGGCAGGA TGGGGACCCT GCTTTCCTCT
1101 ACTTGCTGTC AGCTCCTCGA GAAGCCCCAG CCACAGGACC TAGCCCTCAG
1151 CACCCCCAGA AGATGGACGG GGAACTTGGA CGCTTGTTTC CCCCATCATT
1201 GGGGCTACCC CCAGGCCCCC AGCCAGCTGC CTCCAGCCTG CCCAGTCCAC
1251 TCCAGCCCAG CTGGTCCTGT CCTTCCTGCA CCTTCATCAA TGCCCCAGAC
1301 CGCCCTGGCT GTGAGATGTG TAGCACCCAG AGGCCCTGCA CTTGGGACCC
1351 CCTTGCTGCA GCTTCCACCT AGCAGCCACC AGAGGTTACA AGGGGAGAGT
1401 GGCCCTTCCC TCACAAGTCC GACATCTCCA GGCCCCCACT GAACTCCGGG
1451 GACCTCTACT GACTGCTTGC TGGGACAGTC ACCAGGGTTG GGGGGAAGGG
1501 CCACAAAATG AAACCATTAA AGACCCTTAA GAGCCAAAAA AAAAAAAAAA
1551 AAAAAAAAAA AAAAAAAAAA AAAAAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 209 bp to 1369 bp; peptide length: 387

```
Category: similarity to known protein
Classification: Cell signaling/communication


  1 MAPPAGGAAA AASDLGSAAV LLAVHAAVRP LGAGPDAEAQ LRRLQLSADP
 51 ERPGRFRLEL LGAGPGAVNL EWPLESVSYT IRGPTQHELQ PPPGGPGTLS
101 LHFLNPQEAQ RWAVLVRGAT VEGQNGSKSN SPPALGPEAC PVSLPSPPEA
151 STLKGPPPEA DLPRSPGNLT EREELAGSLA RAIAGGDEKG AAQVAAVLAQ
201 HRVALSVQLQ EACFPPGPIR LQVTLEDAAS AASAASSAHV ALQVHPHCTV
251 AALQEQVFSE LGFPPAVQRW VIGRCLCVPE RSLASYGVRQ DGDPAFLYLL
301 SAPREAPATG PSPQHPQKMD GELGRLFPPS LGLPPGPQPA ASSLPSPLQP
351 SWSCPSCTFI NAPDRPGCEM CSTQRPCTWD PLAAAST

                         BLASTP hits

No BLASTP hits available

          Alert BLASTP hits for DKFZphtes3_23n19, frame 2

PIR:JC5983 protein kinase C-interacting RBCC protein 1 - rat, N = 1,
Score = 353, P = 2.8e-32

TREMBL:AB011369_1 product: "RBCK2"; Rattus norvegicus mRNA for RBCK2,
complete cds., N = 1, Score = 353, P = 2.8e-32

TREMBL:U67322_1 gene: "XAP4"; product: "HBV associated factor"; Human
HBV associated factor (XAP4) mRNA, complete cds., N = 1, Score = 286, P
= 8.5e-25

TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus
musculus UbcM4 interacting protein 28 mRNA, complete cds., N = 1, Score
= 367, P = 9.3e-34


>TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
     UbcM4 interacting protein 28 mRNA, complete cds.
            Length = 498

  HSPs:

 Score = 367 (55.1 bits), Expect = 9.3e-34, P = 9.3e-34
 Identities = 95/212 (44%), Positives = 129/212 (60%)

Query:   175 LAGSLARAIAGGDEKGAAQVAAVLAQHRVALSVQLQEACFPPGPIRLQVTLEDAASAASA 234
             +A SLARA+AGGDE+ A + A  LA+ RV L VQ++     P   IRL V++EDA
Sbjct:     1 MALSLARAVAGGDEQAAIKYATWLAEQRVPLRVQVKPEVSPTQDIRLCVSVEDAYM---- 56

Query:   235 ASSAHVALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGDP 294
              +  + L V P  TVA+L++ VF + GFPP++Q+WV+G+ L    + +L S+G+R++GD
Sbjct:    57 -HTVTIWLTVRPDMTVASLKDMVFLDYGFPPSLQQWVVGQRLARDQETLHSHGIRRNGDG 115

Query:   295 AFLYLLSAPREAPATGPSPQHPQK-----MDGELG--RLFPPSLG-LPPG-PQPAASSLP 345
             A+LYLLSA       T  +PQ  Q+      M  +LG   L   S G L P  P+P  +  P
Sbjct:   116 AYLYLLSARN----TSLNPQELQRQRQLRMLEDLGFKDLTLQSRGPLEPVLPKPRTNQEP 171

Query:   346 -----SPLQP--SWSCPSCTFINAPDRPGCEMCSTQRPCTW 379
                  +P  P   W CP CTFIN P RPGCEMC   RP T+
Sbjct:   172 GQPDAAPESPPVGWQCPGCTFINKPTRPGCEMCCRARPETY 212


          Pedant information for DKFZphtes3_23n19, frame 2
          ------------------------------------------------

                  Report for DKFZphtes3_23n19.2

[LENGTH]        387
[MW]            39949.29
[pI]            5.53
[HOMOL]         TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
UbcM4 interacting protein 28 mRNA, complete cds. 1e-22
[BLOCKS]        BL00578B
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY    17.57 %


SEQ    MAPPAGGAAAAASDLGSAAVLLAVHAAVRPLGAGPDAEAQLRRLQLSADPERPGRFRLEL
SEG    .xxxxxxxxxxxxxxxxxxxxxxxxxx.................................
PRD    cccccchhhhhhhhhhhhhhhhhhhhhhcccccccccchhhhhhhhhhhcccccccccceeee
```

```
SEQ    LGAGPGAVNLEWPLESVSYTIRGPTQHELQPPPGGPGTLSLHFLNPQEAQRWAVLVRGAT
SEG    ............................................................
PRD    cccccccceeeccccceeeeeecccccccccccccccccceeeeeeccccchhhhhheeeecce

SEQ    VEGQNGSKSNSPPALGPEACPVSLPSPPEASTLKGPPPEADLPRSPGNLTEREELAGSLA
SEG    ............................................................
PRD    eeccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhh

SEQ    RAIAGGDEKGAAQVAAVLAQHRVALSVQLQEACFPPGPIRLQVTLEDAASAASAASSAHV
SEG    .................................................xxxxxxxxxxx..
PRD    hhhhccccchhhhhhhhhhhhhhhhhhcccccccccccccceeeccchhhhhhhhhhhhhee

SEQ    ALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGDPAFLYLL
SEG    ............................................................
PRD    eeeccccchhhhhhhhhhhhccccccchhhhhhhhhhhcccccccccccccccceeeeec

SEQ    SAPREAPATGPSPQHPQKMDGELGRLFPPSLGLPPGPQPAASSLPSPLQPSWSCPSCTFI
SEG    ..........................xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx...
PRD    ccccccccchhhhhhhhhhhhcccccccccccccccccccccccccccccccccccceee

SEQ    NAPDRPGCEMCSTQRPCTWDPLAAAST
SEG    ...........................
PRD    ccccccccccccccccccccceeeccc
```

(No Prosite data available for DKFZphtes3_23n19.2)

(No Pfam data available for DKFZphtes3_23n19.2)

similarity to rat protein kinase C-interacting RBCC protein 1

start at Bp 209 matches kozak consensus PyNNatgG
similarity to of C-terminal part to N-terminus of RBCK1

Sequenced by LMU

Locus: unknown

Insert length: 1579 bp
Poly A stretch at pos. 1535, polyadenylation signal at pos. 1515

```
   1 CGGAGACCCT CGGGCCGTGT CCATTTGTGG GCAAAGCCAG CGGGGCAGGC
  51 TTGGCCAGAG TGCACCACTC GGCGCCGTCC CAGGCCCGAC GCTCTGGGCG
 101 CGCCCGGAAC CCCAGGTTCG CGGCCCGTGT TTCCGACCGG CGGAGGGGGC
 151 TCAGCGGCCC GATCCCACGG AAGCGCGCTC GGAGGGGTGG GACCCGGCCG
 201 GACCGGAGAT GGCGCCGCCA GCGGGCGGGG CGGCGGCGGC GGCCTCGGAC
 251 TTGGGCTCCG CCGCAGTGCT CTTGGCTGTG CACGCCGCGG TGAGGCCGCT
 301 GGGCGCCGGG CCAGACGCCG AGGCACAGCT GCGGAGGCTG CAGCTGAGCG
 351 CGGACCCTGA GAGGCCTGGG CGCTTCCGGC TGGAGCTGCT GGGCGCGGGA
 401 CCTGGGGCGG TTAATTTGGA GTGGCCCCTG GAGTCAGTTT CCTACACCAT
 451 CCGAGGCCCC ACCCAGCACG AGCTACAGCC TCCACCAGGA GGGCCTGGAA
 501 CCCTCAGCCT GCACTTCCTC AACCCTCAGG AAGCTCAGCG GTGGGCAGTC
 551 CTAGTCCGAG GTGCCACCGT GGAAGGACAG AATGGCAGCA AGAGCAACTC
 601 ACCACCAGCC TTGGGCCCAG AAGCATGCCC TGTCTCCCTG CCCAGTCCCC
 651 CGGAAGCCTC CACACTCAAG GGCCCTCCAC CTGAGGCAGA TCTTCCTAGG
 701 AGCCCTGGAA ACTTGACGGA GAGAGAAGAG CTGGCAGGGA GCCTGGCCCG
 751 GGCTATTGCA GGTGGAGACG AGAAGGGGGC AGCCCAAGTG GCAGCCGTCC
 801 TGGCCCAGCA TCGTGTGGCC CTGAGTGTTC AGCTTCAGGA GGCCTGCTTC
 851 CCACCTGGCC CCATCAGGCT GCAGGTCACA CTTGAAGACG CTGCCTCTGC
 901 CGCATCCGCC GCGTCCTCTG CACACGTTGC CCTGCAGGTC CACCCCCACT
 951 GCACTGTTGC AGCTCTCCAG GAGCAGGTGT TCTCAGAGCT CGGTTTCCCG
1001 CCAGCCGTGC AACGCTGGGT CATCGGACGG TGCCTGTGTG TGCCTGAGCG
1051 CAGCCTTGCC TCTTACGGGG TTCGGCAGGA TGGGGACCCT GCTTTCCTCT
1101 ACTTGCTGTC AGCTCCTCGA GAAGCCCCAG CCACAGGACC TAGCCCTCAG
1151 CACCCCCAGA AGATGGACGG GGAACTTGGA CGCTTGTTTC CCCCATCATT
1201 GGGGCTACCC CCAGGCCCCC AGCCAGCTGC CTCCAGCCTG CCCAGTCCAC
1251 TCCAGCCCAG CTGGTCCTGT CCTTCCTGCA CCTTCATCAA TGCCCCAGAC
1301 CGCCCTGGCT GTGAGATGTG TAGCACCCAG AGGCCCTGCA CTTGGGACCC
1351 CCTTGCTGCA GCTTCCACCT AGCAGCCACC AGAGGTTACA AGGGGAGAGT
1401 GGCCCTTCCC TCACAAGTCC GACATCTCCA GGCCCCCACT GAACTCCGGG
1451 GACCTCTACT GACTGCTTGC TGGGACAGTC ACCAGGGTTG GGGGGAAGGG
1501 CCACAAAATG AAACCATTAA AGACCCTTAA GAGCCAAAAA AAAAAAAAAA
1551 AAAAAAAAAA AAAAAAAAAA AAAAAAAAG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 209 bp to 1369 bp; peptide length: 387
Category: similarity to known protein
Classification: Cell signaling/communication

```
  1 MAPPAGGAAA AASDLGSAAV LLAVHAAVRP LGAGPDAEAQ LRRLQLSADP
 51 ERPGRFRLEL LGAGPGAVNL EWPLESVSYT IRGPTQHELQ PPPGGPGTLS
101 LHFLNPQEAQ RWAVLVRGAT VEGQNGSKSN SPPALGPEAC PVSLPSPPEA
151 STLKGPPPEA DLPRSPGNLT EREELAGSLA RAIAGGDEKG AAQVAAVLAQ
201 HRVALSVQLQ EACFPPGPIR LQVTLEDAAS AASAASSAHV ALQVHPHCTV
251 AALQEQVFSE LGFPPAVQRW VIGRCLCVPE RSLASYGVRQ DGDPAFLYLL
301 SAPREAPATG PSPQHPQKMD GELGRLFPPS LGLPPGPQPA ASSLPSPLQP
351 SWSCPSCTFI NAPDRPGCEM CSTQRPCTWD PLAAAST
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_23n19, frame 2

PIR:JC5983 protein kinase C-interacting RBCC protein 1 - rat, N = 1, Score = 353, P = 2.8e-32

TREMBL:AB011369_1 product: "RBCK2"; Rattus norvegicus mRNA for RBCK2, complete cds., N = 1, Score = 353, P = 2.8e-32

TREMBL:U67322_1 gene: "XAP4"; product: "HBV associated factor"; Human HBV associated factor (XAP4) mRNA, complete cds., N = 1, Score = 286, P = 8.5e-25

TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus UbcM4 interacting protein 28 mRNA, complete cds., N = 1, Score = 367, P = 9.3e-34

```
>TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
     UbcM4 interacting protein 28 mRNA, complete cds.
          Length = 498

  HSPs:

 Score = 367 (55.1 bits), Expect = 9.3e-34, P = 9.3e-34
 Identities = 95/212 (44%), Positives = 129/212 (60%)

Query:   175 LAGSLARAIAGGDEKGAAQVAAVLAQHRVALSVQLQEACFPPGPIRLQVTLEDAASAASA 234
             +A SLARA+AGGDE+ A + A  LA+ RV L VQ++     P   IRL V++EDA
Sbjct:     1 MALSLARAVAGGDEQAAIKYATWLAEQRVPLRVQVKPEVSPTQDIRLCVSVEDAYM---- 56

Query:   235 ASSAHVALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGDP 294
               +   + L V P  TVA+L++ VF + GFPP++Q+WV+G+ L    + +L S+G+R++GD
Sbjct:    57 -HTVTIWLTVRPDMTVASLKDMVFLDYGFPPSLQQWVVGQRLARDQETLHSHGIRRNGDG 115

Query:   295 AFLYLLSAPREAPATGPSPQHPQK-----MDGELG--RLFPPSLG-LPPG-PQPAASSLP 345
             A+LYLLSA       T  +PQ  Q+       M  +LG   L   S G L P  P+P  +  P
Sbjct:   116 AYLYLLSARN----TSLNPQELQRQRQLRMLEDLGFKDLTLQSRGPLEPVLPKPRTNQEP 171

Query:   346 -----SPLQP--SWSCPSCTFINAPDRPGCEMCSTQRPCTW 379
                  +P  P   W CP CTFIN P RPGCEMC   RP T+
Sbjct:   172 GQPDAAPESPPVGWQCPGCTFINKPTRPGCEMCCRARPETY 212
```

Pedant information for DKFZphtes3_23n19, frame 2

Report for DKFZphtes3_23n19.2

```
[LENGTH]        387
[MW]            39949.29
[pI]            5.53
[HOMOL]         TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
UbcM4 interacting protein 28 mRNA, complete cds. 1e-22
[BLOCKS]        BL00578B
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY    17.57 %

SEQ    MAPPAGGAAAAASDLGSAAVLLAVHAAVRPLGAGPDAEAQLRRLQLSADPERPGRFRLEL
SEG    .xxxxxxxxxxxxxxxxxxxxxxxxxxx................................
PRD    ccccchhhhhhhhhhhhhhhhhhhhccccccccccchhhhhhhhhhhcccccccccceeee

SEQ    LGAGPGAVNLEWPLESVSYTIRGPTQHELQPPPGGPGTLSLHFLNPQEAQRWAVLVRGAT
SEG    ............................................................
PRD    ccccccceeecccceeeeeecccccccccccccccccceeeeeeccchhhhhheeeecce

SEQ    VEGQNGSKSNSPPALGPEACPVSLPSPPEASTLKGPPPEADLPRSPGNLTEREELAGSLA
SEG    ............................................................
PRD    eecccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhh

SEQ    RAIAGGDEKGAAQVAAVLAQHRVALSVQLQEACFPPGPIRLQVTLEDAASAASAASSAHV
SEG    ...............................................xxxxxxxxxxx..
PRD    hhhhcccchhhhhhhhhhhhhhhhhhccccccccccccccceeecchhhhhhhhhhhhee

SEQ    ALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGDPAFLYLL
SEG    ............................................................
PRD    eeecccccchhhhhhhhhhhhccccccchhhhhhhhhhhhcccccccccccccccceeeeec

SEQ    SAPREAPATGPSPQHPQKMDGELGRLFPPSLGLPPGPQPAASSLPSPLQPSWSCPSCTFI
SEG    ...........................xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx...
PRD    ccccccccchhhhhhhhhhhhcccccccccccccccccccccccccccccccccccceee

SEQ    NAPDRPGCEMCSTQRPCTWDPLAAAST
SEG    ...........................
PRD    ccccccccccccccccccccceeeccc
```

(No Prosite data available for DKFZphtes3_23n19.2)

(No Pfam data available for DKFZphtes3_23n19.2)

DKFZphtes3_26g22
----

group: intracellular transport/trafficking

DKFZphtes3_26g22 encodes a novel 898 amino acid protein with similarity to kinesins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a kinesin motor domain signature. Kinesin is a microtubule-associated force-producing protein that play a role in organelle transport. It is an oligomeric complex composed of two heavy chains and two light chains. The kinesin motor activity is directed toward the microtubule's plus end. The heavy chain contains a large globular N-terminal domain which is responsible for the motor activity of kinesin, which is known to hydrolyze ATP and to bind and move on microtubules. Several proteins involved in chromosome segregation and cell divsion contain this motor domain, such as drosophila claret segregational protein (ncd), Drosophila kinesin-like protein (nod), human CENP-E and human mitotic kinesin-like protein-1 (MKLP-1). The novel protein is a new kinesin like proptein.

The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division.

strong similarity to kinesins

Sequenced by EMBL

Locus: unknown

Insert length: 3032 bp
No poly A stretch found, no polyadenylation signal found

```
   1 CTGAAGCGCT GGGAGGCGGA CATTAAAGTG AAGTGGTTGC GGTAACCTGG
  51 CCTGGGCCTG AAGTGAGTGA GAGGCACATG AAGAGAAGTA TTCAAGTATT
 101 TATACAGATA GGAATCAAGA TAATCAACAA TGTCTGTCAC TGAGGAAGAC
 151 CTGTGCCACC ATATGAAAGT AGTAGTTCGT GTACGTCCGG AAAACACTAA
 201 AGAAAAAGCA GCTGGATTTC ATAAAGTGGT TCATGTTGTG GATAAACATA
 251 TCCTAGTTTT TGATCCCAAA CAAGAAGAAG TCAGTTTTTT CCATGGAAAG
 301 AAAACTACAA ATCAAAATGT TATAAAGAAA CAAAATAAGG ATCTTAAATT
 351 TGTATTTGAT GCTGTTTTTG ATGAAACGTC AACTCAGTCA GAAGTTTTTG
 401 AACACACTAC TAAGCCAATT CTTCGTAGTT TTTTGAATGG ATATAATTGC
 451 ACAGTACTTG CCTATGGTGC CACTGGTGCT GGGAAGACCC ACACTATGCT
 501 AGGATCAGCT GATGAACCTG GAGTGATGTA TCTAACAATG TTACACCTTT
 551 ACAAATGCAT GGATGAGATT AAAGAAGAGA AAATATGTAG TACTGCAGTT
 601 TCATATCTGG AGGTATATAA TGAACAGATT CGTGATCTCT TAGTAAATTC
 651 AGGGCCACTT GCTGTCCGGG AAGATACCCA AAAAGGGGTG GTCGTTCATG
 701 GACTTACTTT ACACCAGCCC AAATCCTCAG AAGAAATTTT ACATTTATTG
 751 GATAATGGAA ACAAAAACAG GACACAACAT CCCACTGATA TGAATGCCAC
 801 ATCTTCTCGT TCTCATGCTG TTTTCCAAAT TTACTTGCGA CAACAAGACA
 851 AAACAGCAAG TATCAATCAA AATGTCCGTA TTGCCAAGAT GTCACTCATT
 901 GACCTGGCAG GATCTGAGCG AGCAAGTACT TCCGGTGCTA AGGGGACCCG
 951 ATTTGTAGAA GGCACAAATA TTAATAGATC ACTTTTAGCT CTTGGGAATG
1001 TCATCAATGC CTTAGCAGAT TCAAAGAGAA AGAATCAGCA TATCCCTTAC
1051 AGAAATAGTA AGCTTACTCG CTTGTTAAAG GATTCTCTTG GAGGAAACTG
1101 TCAAACTATA ATGATAGCTG CTGTTAGTCC TTCCTCTGTA TTCTACGATG
1151 ACACATATAA CACTCTTAAG TATGCTAACC GGGCAAAGGA CATTAAATCT
1201 TCTTTGAAGA GCAATGTTCT TAATGTCAAT AATCATATAA CTCAATATGT
1251 AAAGATCTGT AATGAGCAGA AGGCAGAGAT TTTATTGTTA AAAGAAAAAC
1301 TAAAAGCCTA TGAAGAACAG AAAGCCTTCA CTAATGAAAA TGACCAAGCA
1351 AAGTTAATGA TTTCAAACCC TCAGGAAAAA GAAATCGAAA GGTTTCAAGA
1401 AATCCTGAAC TGCTTGTTCC AGAATCGAGA AGAAATTAGA CAAGAATATC
1451 TGAAGTTGGA AATGTTACTT AAAGAAAATG AACTTAAATC ATTCTACCAA
1501 CAACAGTGCC ATAAACAAAT AGAAATGATG TGTTCTGAAG ACAAAGTAGA
1551 AAAGGCCACT GGAAAACGAG ATCATAGACT TGCAATGTTG AAAACTCGTC
1601 GCTCCTACCT GGAGAAAAGG AGGGAGGAGG AATTGAAGCA ATTTGATGAG
1651 AATACTAATT GGCTCCATCG TGTCGAAAAA GAAATGGGAC TCTTAAGTCA
1701 AAACGGTCAT ATTCCAAAGG AACTCAAGAA AGATCTTCAT TGTCACCATT
1751 TGCACCTCCA GAACAAAGAT TTGAAAGCAC AAATTAGACA TATGATGGAT
1801 CTAGCTTGTC TTCAGGAACA GCAACACAGG CAGACTGAAG CAGTATTGAA
1851 TGCTTTACTT CCAACCCTAA GAAAACAATA TTGCACATTA AAAGAAGCCG
1901 GCCTGTCAAA TGCTGCTTTT GAATCTGACT TCAAAGAGAT CGAACATTTG
1951 GTAGAGAGGA AAAAAGTGGT AGTTTGGGCT GACCAAACTG CCGAACAACC
2001 AAAGCAAAAC GATCTACCAG GGATTTCTGT TCTTATGACC TTTCCACAAC
2051 TTGGACCAGT TCAGCCTATT CCTTGTTGCT CATCTTCAGG TGGAACTAAT
2101 CTGGTTAAGA TTCCTACAGA AAAAAGAACT CGGAGAAAAC TAATGCCATC
2151 TCCCTTGAAA GGACAGCATA CTCTAAAGTC TCCACCATCT CAAAGTGTGC
2201 AGCTCAATGA TTCTCTTAGC AAAGAACTTC AGCCTATTGT ATATACACCA
2251 GAAGACTGTA GAAAAGCTTT TCAAAATCCG TCTACAGTAA CCTTAATGAA
2301 ACCATCATCA TTTACTACAA GTTTTCAGGC TATCAGCTCA AACATAAACA
2351 GTGATAATTG TCTGAAAATG TTGTGTGAAG TAGCTATCCC TCATAATAGA
```

```
2401 AGAAAAGAAT GTGGACAGGA GGACTTGGAC TCTACATTTA CTATATGTGA
2451 AGACATCAAG AGCTCGAAGT GTAAATTACC CGAACAAGAA TCACTACCAA
2501 ATGATAACAA AGACATTTTA CAACGGCTTG ATCCTTCTTC ATTCTCAACT
2551 AAGCATTCTA TGCCTGTACC AAGCATGGTG CCATCCTACA TGGCAATGAC
2601 TACTGCTGCC AAAAGGAAAC GGAAATTAAC AAGTTCTACA TCAAACAGTT
2651 CGTTAACTGC AGACGTAAAT TCTGGATTTG CCAAACGTGT TCGACAAGAT
2701 AATTCAAGTG AGAAGCACTT ACAAGAAAAC AAACCAACAA TGGAACATAA
2751 AAGAAACATC TGTAAAATAA ATCCAAGCAT GGTTAGAAAA TTTGGAAGAA
2801 ATATTTCAAA AGGAAATCTA AGATAAATCA CTTCAAAACC AAGCAAAATG
2851 AAGTTGATCA AATCTGCTTT TCAAAGTTTA TCAATACCCT TTCAAAAATA
2901 TATTTAAAAT CTTTGAAAGA AGACCCATCT TAAAGCTAAG TTTACCCAAG
2951 TACTTTCAGC AAGCAGAAAA ATGAAACTCT TTGTTTTCTT CTTTTGTGTT
3001 CTAAAAAAAT AAAATTTCAA AAGAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 130 bp to 2823 bp; peptide length: 898
Category: strong similarity to known protein
Classification: Cell structure/motility
Prosite motifs: ATP_GTP_A (113-121)
KINESIN_MOTOR_DOMAIN1 (252-264)

```
  1 MSVTEEDLCH HMKVVVRVRP ENTKEKAAGF HKVVHVVDKH ILVFDPKQEE
 51 VSFFHGKKTT NQNVIKKQNK DLKFVFDAVF DETSTQSEVF EHTTKPILRS
101 FLNGYNCTVL AYGATGAGKT HTMLGSADEP GVMYLTMLHL YKCMDEIKEE
151 KICSTAVSYL EVYNEQIRDL LVNSGPLAVR EDTQKGVVVH GLTLHQPKSS
201 EEILHLLDNG NKNRTQHPTD MNATSSRSHA VFQIYLRQQD KTASINQNVR
251 IAKMSLIDLA GSERASTSGA KGTRFVEGTN INRSLLALGN VINALADSKR
301 KNQHIPYRNS KLTRLLKDSL GGNCQTIMIA AVSPSSVFYD DTYNTLKYAN
351 RAKDIKSSLK SNVLNVNNHI TQYVKICNEQ KAEILLLKEK LKAYEEQKAF
401 TNENDQAKLM ISNPQEKEIE RFQEILNCLF QNREEIRQEY LKLEMLLKEN
451 ELKSFYQQQC HKQIEMMCSE DKVEKATGKR DHRLAMLKTR RSYLEKRREE
501 ELKQFDENTN WLHRVEKEMG LLSQNGHIPK ELKKDLHCHH LHLQNKDLKA
551 QIRHMMDLAC LQEQQHRQTE AVLNALLPTL RKQYCTLKEA GLSNAAFESD
601 FKEIEHLVER KKVVVWADQT AEQPKQNDLP GISVLMTFPQ LGPVQPIPCC
651 SSSGGTNLVK IPTEKRTRRK LMPSPLKGQH TLKSPPSQSV QLNDSLSKEL
701 QPIVYTPEDC RKAFQNPSTV TLMKPSSFTT SFQAISSNIN SDNCLKMLCE
751 VAIPHNRRKE CGQEDLDSTF TICEDIKSSK CKLPEQESLP NDNKDILQRL
801 DPSSFSTKHS MPVPSMVPSY MAMTTAAKRK RKLTSSTSNS SLTADVNSGF
851 AKRVRQDNSS EKHLQENKPT MEHKRNICKI NPSMVRKFGR NISKGNLR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_26g22, frame 1

SWISSPROT:YB3D_SCHPO PUTATIVE KINESIN-LIKE PROTEIN C2F12.13., N = 3, Score = 874, P = 9e-93

TREMBL:DMU89264_1 product: "kinesin like protein 67a"; Drosophila melanogaster kinesin like protein 67a mRNA, complete cds., N = 1, Score = 880, P = 4.2e-88

TREMBL:SPBC649_1 gene: "SPBC649.01c"; product: "putative kinesin-like protein"; S.pombe chromosome II cosmid c649., N = 3, Score = 814, P = 9.8e-86

PIR:S64238 kinesin-related protein KIP3 - yeast (Saccharomyces cerevisiae), N = 2, Score = 802, P = 2.5e-83

>TREMBL:DMU89264_1 product: "kinesin like protein 67a"; Drosophila

melanogaster kinesin like protein 67a mRNA, complete cds.
Length = 814

HSPs:

Score = 880 (132.0 bits), Expect = 4.2e-88, P = 4.2e-88
Identities = 181/345 (52%), Positives = 238/345 (68%)

```
Query:   11 HMKVVVRVRVRPENTKEKAAGFHKVVHVVDKHILVFDPKQEEVSFF-HGKKTTNQNVIKKQN 69
            ++KV VRVRP N +E         ++ V+D+  L+FDP +E+  FF  G K    +++ K+ N
Sbjct:    8 NIKVAVRVRPYNVRELEQKQRSIIKVMDRSALLFDPDEEDDEFFFQGAKQPYRDITKRMN 67

Query:   70 KDLKFVFDAVFDETSTQSEVFEHTTKPILRSFLNGYNCTVLAYGATGAGKTHTMLGSADE 129
            K L   FD VFD  ++  ++FE  T P++ + LNGYNC+V  YGATGAGKT TMLGS
Sbjct:   68 KKLTMEFDRVFDIDNSNQDLFEECTAPLVDAVLNGYNCSVFVYGATGAGKTFTMLGSEAH 127

Query:  130 PGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVREDTQKGVVV 189
            PG+ YLTM  L+  +     + +      VSYLEVYNE + +LL  SGPL +RED   GVVV
Sbjct:  128 PGLTYLTMQDLFDKIQAQSDVRKFDVGVSYLEVYNEHVMNLLTKSGPLKLREDNN-GVVV 186

Query:  190 HGLTLHQPKSSEEILHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQDKTASINQNV 249
            GL L    S+EE+L +L  GN +RTQHPTD NA  SSRSHA+FQ+++R  ++     + V
Sbjct:  187 SGLCLTPIYSAEELLRMLMLGNSHRTQHPTDANAESSRSHAIFQVHIRITERKTDTKRTV 246

Query:  250 RIAKMSLIDLAGSERASTSGAKGTRFVEGTNINRSLLALGNVINALADSKRKNQHIPYRN 309
               K+S+IDLAGSERA+++   G RF EG +IN+SLLALGN IN LAD  +   HIPYR+
Sbjct:  247 ---KLSMIDLAGSERAASTKGIGVRFKEGASINKSLLALGNCINKLADGLK---HIPYRD 300

Query:  310 SKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDI 355
            S LTR+LKDSLGGNC+T+M+A VS SS+ Y+DTYNTLKYA+RAK I
Sbjct:  301 SNLTRILKDSLGGNCRTLMVANVSMSSLTYEDTYNTLKYASRAKKI 346
```

Pedant information for DKFZphtes3_26g22, frame 1

Report for DKFZphtes3_26g22.1

```
[LENGTH]     898
[MW]         102281.63
[pI]         9.09
[HOMOL]      SWISSPROT:YB3D_SCHPO PUTATIVE KINESIN-LIKE PROTEIN C2F12.13. 3e-97
[FUNCAT]     30.04 organization of cytoskeleton   [S. cerevisiae, YGL216w] 2e-88
[FUNCAT]     03.22 cell cycle control and mitosis [S. cerevisiae, YGL216w] 2e-88
[FUNCAT]     08.22 cytoskeleton-dependent transport     [S. cerevisiae, YGL216w] 2e-88
[FUNCAT]     30.10 nuclear organization     [S. cerevisiae, YGL216w] 2e-88
[FUNCAT]     09.10 nuclear biogenesis     [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]     06.10 assembly of protein complexes   [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]     03.13 meiosis  [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]     11.01 stress response  [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]     03.07 pheromone response, mating-type determination, sex-specific proteins
     [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]     30.05 organization of centrosome     [S. cerevisiae, YPR141c] 5e-42
[FUNCAT]     03.04 budding, cell polarity and filament formation [S. cerevisiae, YKL079w]
4e-28
[BLOCKS]     BL00411H
[BLOCKS]     BL00411G
[BLOCKS]     BL00411F
[BLOCKS]     BL00411E Kinesin motor domain proteins
[BLOCKS]     BL00411C Kinesin motor domain proteins
[BLOCKS]     BL00411B Kinesin motor domain proteins
[BLOCKS]     BL00411A Kinesin motor domain proteins
[SCOP]       d2kin.1 3.29.1.5.3 Kinesin  [Rat (Rattus norvegicus)  1e-117
[SCOP]       d3kar__ 3.29.1.5.4 Kinesin  [Baker's yeast (Saccharomyce 1e-112
[PIRKW]      nucleus 6e-87
[PIRKW]      heterodimer 4e-68
[PIRKW]      DNA binding 9e-60
[PIRKW]      heterotetramer 2e-54
[PIRKW]      mitosis 9e-60
[PIRKW]      microtubule binding 4e-68
[PIRKW]      ATP 6e-87
[PIRKW]      phosphoprotein 5e-59
[PIRKW]      heterotrimer 4e-68
[PIRKW]      purine nucleotide binding 1e-26
[PIRKW]      P-loop 6e-87
[PIRKW]      coiled coil 4e-68
[PIRKW]      heptad repeat 3e-62
[PIRKW]      methylated amino acid 2e-54
[PIRKW]      hydrolase 2e-54
[PIRKW]      GTP binding 1e-60
```

```
[PIRKW]       cell division 5e-57
[SUPFAM]      kinesin-related protein KIP1 3e-50
[SUPFAM]      kinesin-related protein CIN8 7e-33
[SUPFAM]      kinesin heavy chain 2e-54
[SUPFAM]      suppressor protein SMY1 1e-26
[SUPFAM]      kinesin-related protein KIF3 4e-68
[SUPFAM]      kinesin-related protein KIF2 1e-46
[SUPFAM]      kinesin-related protein unc-104 7e-60
[SUPFAM]      unassigned kinesin-related proteins 6e-87
[SUPFAM]      centromere protein E 3e-54
[SUPFAM]      kinesin-related protein KLP61F 5e-57
[SUPFAM]      kinesin-related protein MKLP-1 2e-28
[SUPFAM]      pleckstrin repeat homology 7e-60
[SUPFAM]      kinesin-related protein KIF1B 4e-61
[SUPFAM]      kinesin motor domain homology 6e-87
[SUPFAM]      kinesin-related protein KLPA 1e-43
[SUPFAM]      kinesin-related protein nodA 1e-30
[SUPFAM]      kinesin-related protein Eg5 5e-59
[PROSITE]     ATP_GTP_A      1
[PROSITE]     KINESIN_MOTOR_DOMAIN1  1
[PFAM]        Kinesin motor domain
[KW]          Irregular
[KW]          3D
[KW]          LOW_COMPLEXITY     8.57 %
```

```
SEQ    MSVTEEDLCHHMKVVVRVRPENTKEKAAGFHKVVHVVDKHILVFDPKQEEVSFFHGKKTT
SEG    ............................................................
3kar-  .......................................................TBEEE

SEQ    NQNVIKKQNKDLKFVFDAVFDETSTQSEVFEHTTKPILRSFLNGYNCTVLAYGATGAGKT
SEG    ............................................................
3kar-  EEEETTTTTTEEEEEETEEETTTTCHHHHHHHHH-HHHGGGCCCEEEEEECTTTTCHH

SEQ    HTMLGSADEPGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVR
SEG    ............................................................
3kar-  HHHHTTTT--THHHHHHHHHHHHHHHGGGCEEEEEEEEEEEETTEEEETT-TCCCCEEE

SEQ    EDTQKGVVVHGLTLHQPKSSEEILHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQD
SEG    ............................................................
3kar-  EETTTEEEEETTCCEEECCGGGHHHHHHHHHHHHCCTTTTCHHHHHHCEEEEEEEEEEEE

SEQ    KTASINQNVRIAKMSLIDLAGSERASTSGAKGTRFVEGTNINRSLLALGNVINALADSKR
SEG    ............................................................
3kar-  TTTTCEE---EEEEEEEECCCCCCCCCC----HHHHHHHHHHHHHHHHHHHHHHHHTTTT

SEQ    KNQHIPYRNSKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIKSSLK
SEG    .......................................................xxxxx
3kar-  TTTCCTTTTTHHHHHHGGGCTTTTEEEEEEEECCCGGGHHHHHHHHHHHHH.........

SEQ    SNVLNVNNHITQYVKICNEQKAEILLLKEKLKAYEEQKAFTNENDQAKLMISNPQEKEIE
SEG    xxxxxxxx..........xxxxxxxxxxxxxxxxxxxxx.....................
3kar-  ............................................................

SEQ    RFQEILNCLFQNREEIRQEYLKLEMLLKENELKSFYQQQCHKQIEMMCSEDKVEKATGKR
SEG    ...................xxxxxxxxxxxxx............................
3kar-  ............................................................

SEQ    DHRLAMLKTRRSYLEKRREEELKQFDENTNWLHRVEKEMGLLSQNGHIPKELKKDLHCHH
SEG    .................................................xxxxxxxxxxx
3kar-  ............................................................

SEQ    LHLQNKDLKAQIRHMMDLACLQEQQHRQTEAVLNALLPTLRKQYCTLKEAGLSNAAFESD
SEG    xxx.........................................................
3kar-  ............................................................

SEQ    FKEIEHLVERKKVVVWADQTAEQPKQNDLPGISVLMTFPQLGPVQPIPCCSSSGGTNLVK
SEG    ............................................................
3kar-  ............................................................

SEQ    IPTEKRTRRKLMPSPLKGQHTLKSPPSQSVQLNDSLSKELQPIVYTPEDCRKAFQNPSTV
SEG    ............................................................
3kar-  ............................................................

SEQ    TLMKPSSFTTSFQAISSNINSDNCLKMLCEVAIPHNRRKECGQEDLDSTFTICEDIKSSK
SEG    ............................................................
3kar-  ............................................................

SEQ    CKLPEQESLPNDNKDILQRLDPSSFSTKHSMPVPSMVPSYMAMTTAAKRKRKLTSSTSNS
SEG    ...............................................xxxxxxxxxxxxx
3kar-  ............................................................
```

```
SEQ     SLTADVNSGFAKRVRQDNSSEKHLQENKPTMEHKRNICKINPSMVRKFGRNISKGNLR
SEG     xxx.......................................................
3kar-   ..........................................................
```

Prosite for DKFZphtes3_26g22.1

```
PS00017     113->121    ATP_GTP_A              PDOC00017
PS00411     252->264    KINESIN_MOTOR_DOMAIN1  PDOC00343
```

Pfam for DKFZphtes3_26g22.1

```
HMM_NAME     Kinesin motor domain

HMM               *RCRPlNeREindgcscvVQWPpWtGyktvhnghegds...........
                    R+RP N +E+++G  +VV + + +    +  +++E  S
Query         17   RVRPENTKEKAAGFHKVVHVVD-KHILVFDPKQEEVSFFHGKKTTNQNV    64

HMM               .....phksFtFDHVFWWncTQedVYdtvAHPIVDDcFhGYNCTIFAYGQ
                       +     F+FD VF+ ++TQ +V++ +   PI+ ++++GYNCT++AYG
Query         65  IKKQNKDLKFVFDAVFDETSTQSEVFEHTTKPILRSFLNGYNCTVLAYGA   114

HMM               TGSGKTYTMMGpggehPDHmGIIPRcCHDIFdrIdkfqekDhdFWhVkCS
                  TG+GKT+TM G     + D+ G+    + +++++  D     + +    + +S
Query        115  TGAGKTHTMLG----SADEPGVMYLTMLHLYKCMDEIK-EEKIC-STAVS   158

HMM               YMEIYNEeIYDLLCPnPqhMkpLnIHEHPNMGpYVqGCTEfHVcSYeDac
                  Y+E+YNE+I+DLL+ N    ++PL+++E+    G+ V G+T+    +S E+++
Query        159  YLEVYNEQIRDLLV-N---SGPLAVREDTQKGVVVHGLTLHQPKSSEEIL   204

HMM               hWIWqGnknRHVAaTnMNdhSSRSHtIFTIHVeQrHk..qcdehvcHSKM
                  H+++ GNKNR+   +T MN++SSRSH++F+I ++Q  K  +      V++ KM
Query        205  HLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQDKTASINQNVRIAKM   254

HMM               NLVDLAGSERvnrTGAEGQRlKEGcNINqSLttLGnVInaLaDgqTKYmY
                  +L+DLAGSER++ +GA G+R+ EG+NIN+SL++LGNVINALAD +
Query        255  SLIDLAGSERASTSGAKGTRFVEGTNINRSLLALGNVINALADSK-----   299

HMM               gghgHIPYRDSKLTWlLQDSLGGNcKTcMIACIWPadWNYEETLSTLRYA
                   +++HIPYR SKLT+LL+DSLGGNC T MIA+++P+ + Y++T +TL+YA
Query        300  RKNQHIPYRNSKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYA   349

HMM               dRAKnIkNkPQINEDPcamalWRrYheQIqdMKhqL*
                  +RAK+IK    + N  + +    + +Y +    + K++
Query        350  NRAKDIKSSLKSNVLNVN-NHITQYVKICNEQKAEI    384
```

DKFZphtes3_27d1
----

group: metabolism

DKFZphtes3_27d1 encodes a novel 712 amino acid protein similar to ubiquitin-specific proteases (EC 3.1.2.15).

The novel protein contains both, a ubiquitin carboxyl-terminal hydrolases family 2 signature 1 and signature 2. Pfam predicts a new member of the ubiquitin carboxyl-terminal hydrolases family 2. The ubiquitin system is responsible for the turn over of proteins. Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquinated proteins.
The novel protein is a new member of the ubiquitin carboxyl-terminal hydrolases family 2, represented by proteins such as yeast UBP1-16, human tre-2, human isopeptidase T and others.

The novel protein can find application in modulation of ubiquitin- and protein metabolism in cells.

similarity to ubiquitin-specific proteases

complete cDNA, complete cds, 4 EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2871 bp
Poly A stretch at pos. 2836, no polyadenylation signal found

```
   1 CCAAACCTGA AAGAGGTTGA TTTGTAATGA TTTGCAGGGG GGCACTGGAG
  51 GCAGCGGCCA GGACTTTTCA CTTAGGAGAT CAGCATTTGC CCTGATGGAA
 101 ACTGGGCGAT CCTGCAGGGA CTGACCTCTG AGTTATCCAA AGGCCGACCT
 151 GGGGAAAGAC TGATTTTGAG GTTTTAATAG TTTTCAGATG CTTCAAGTGT
 201 TGTGAACAGA GACTTGTTTG GATTATGCAT TTCTCAGCTA GACTAAATAA
 251 ATGCTAGCAA TGGATACGTG CAAACATGTT GGGCAGCTGC AGCTTGCTCA
 301 AGACCATTCC AGCCTCAACC CTCAGAAATG GCACTGTGTG GACTGCAACA
 351 CGACCGAGTC CATTTGGGCT TGCCTTAGCT GCTCCCATGT TGCCTGTGGA
 401 AGATATATTG AAGAGCATGC ACTCAAGCAC TTTCAAGAAA GCAGTCATCC
 451 TGTTGCATTG GAGGTGAATG AGATGTACGT TTTTTGTTAC CTTTGTGATG
 501 ATTATGTTCT GAATGATAAC GCAACTGGAG ACCTGAAGTT ACTACGACGT
 551 ACATTAAGTG CCATCAAAAG TCAAAATTAT CACTGCACAA CTCGTAGTGG
 601 GAGGTTTTTA CGGTCCATGG GTACAGGTGA TGATTCTTAT TTCTTACATG
 651 ACGGTGCCCA ATCTCTGCTT CAAAGTGAAG ATCAACTGTA TACTGCTCTT
 701 TGGCACAGGA GAAGGATACT AATGGGTAAA ATCTTTCGAA CATGGTTTGA
 751 ACAATCACCC ATTGGAAGAA AAAAGCAAGA AGAACCATTT CAGGAGAAAA
 801 TAGTAGTAAA AAGAGAAGTA AAGAAAAGAC GGCAGGAATT GGAGTATCAA
 851 GTTAAAGCAG AATTGGAAAG TATGCCTCCA AGAAAGAGTT TACGTTTACA
 901 AGGGCTCGCT CAGTCGACCA TAATAGAAAT AGTTTCTGTT CAGGTGCCAG
 951 CACAAACGCC AGCATCACCA GCAAAAGATA AAGTACTCTC TACCTCAGAA
1001 AATGAAATAT CTCAAAAAGT CAGTGACTCC TCAGTTAAAC GAAGGCCAAT
1051 AGTAACTCCT GGTGTAACAG GATTGAGAAA TTTGGGAAAT ACTTGCTATA
1101 TGAATTCTGT TCTTCAGGTG TTGAGTCATT TACTTATTTT TCGACAATGT
1151 TTTTTAAAGC TTGATCTGAA CCAATGGCTG GCTATGACTG CTAGCGAGAA
1201 GACAAGATCT TGTAAGCATC CACCAGTCAC AGATACAGTA GTATATCAAA
1251 TGAATGAATG TCAGGAAAAA GATACAGGTT TTGTTTGCTC CAGACAATCA
1301 AGTCTGTCAT CAGGACTAAG TGGTGGAGCA TCAAAAGGTA GAAAGATGGA
1351 ACTTATTCAG CCAAAGGAGC CAACTTCACA GTACATTTCT CTTTGTCATG
1401 AATTGCATAC TTTGTTCCAA GTCATGTGGT CTGGAAAGTG GGCGTTGGTC
1451 TCACCATTTG CTATGCTACA CTCAGTGTGG AGACTCATTC CTGCCTTTCG
1501 TGGTTACGCC CAACAAGACG CTCAGGAATT TCTTTGTGAA CTTTTAGATA
1551 AAATACAACG TGAATTAGAG ACAACTGGTA CCAGTTTACC AGCTCTTATC
1601 CCCACTTCTC AAAGGAAACT CATCAAACAA GTTCTGAATG TTGTAAATAA
1651 CATTTTTCAT GGACAACTTC TTAGTCAGGT TACATGTCTT GCATGTGACA
1701 ACAAATCAAA TACCATAGAA CCTTTCTGGG ACTTGTCATT GGAGTTTCCA
1751 GAAAGGTATC AATGCAGTGG AAAAGATATT GCTTCCCAGC CATGTCTGGT
1801 TACTGAAATG TTGGCCAAAT TTACAGAAAC TGAAGCTTTA GAAGGAAAAA
1851 TCTACGTATG TGACCAGTGT AACTCAAAGC GTAGAAGGTT TTCCTCCAAA
1901 CCAGTTGTAC TCACAGAAGC CCAGAAACAA CTTATGATAT GCCACCTACC
1951 TCAGGTTCTC AGACTGCACC TCAAACGATT CAGGTGGTCA GGACGTAATA
2001 ACCGAGAGAA GATTGGTGTT CATGTTGGCT TTGAGGAAAT CTTAAACATG
2051 GAGCCCTATT GCTGCAGGGA GACCCTGAAA TCCCTCAGAC CAGAATGCTT
2101 TATCTATGAC TTGTCCGCGG TGGTGATGCA CCATGGGAAA GGATTTGGCT
2151 CAGGGCACTA CACTGCCTAC TGCTATAATT CTGAAGGAGG GTTCTGGGTA
2201 CACTGCAATG ATTCCAAACT AAGCATGTGC ACTATGGATG AAGTATGCAA
2251 GGCTCAAGCT TATATCTTGT TTTATACCCA ACGAGTTACT GAGAATGGAC
```

```
2301 ATTCTAAACT TTTGCCTCCA GAGCTCCTGT TGGGGAGCCA ACATCCCAAT
2351 GAAGACGCTG ATACCTCGTC TAATGAAATC CTTAGCTGAT CCAAAGACAA
2401 TGGGGTTTTC TTCCTGTGAT TTATATATAT ACTTTTTAAA AGACTGATGT
2451 ACCATTTTAA ACTTCATTTT TTCTTGTGAA TCAGTGTATA CTACATTTAT
2501 ACATTTTATA TCTAACAATT TTTTTTTTTT ACAAAGTATA AATGTATATA
2551 TCAACTGAAG GTAACTACTT TTTTCATATT TGGAGTTTTA AACTTTTGGT
2601 GTTTACCTCA GACTGATGTT ACCTCTTTTA TATTTTTATG TCTTAATTGG
2651 CTCGGATGAT GAACTTGTGC AATCTTCTAC CAACAAAGTT CAAGTGGCAT
2701 CATTTTATAT ACATGTATCT TTTTCAGGTA TTTTCTATAC AAATTCTTAA
2751 TAGATGGAAA ATTAGACTCT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AAAAAAAAAA AAAAAAAAAA AAGGGGCGGC CGCTCTAAAA AAAAAAAAAA
2851 AAAAAAAAAA AAAAAAAAAG G
```

BLAST Results

No BLAST result

Medline entries

98072201:
Regulation of ubiquitin-dependent processes by deubiquitinating enzymes.

98431658:
The ubiquitin system.

Peptide information for frame 2

ORF from 251 bp to 2386 bp; peptide length: 712
Category: similarity to known protein
Prosite motifs: UCH_2_1 (274-290)
UCH_2_2 (619-638)
UCH_2_2 (619-638)

```
  1 MLAMDTCKHV GQLQLAQDHS SLNPQKWHCV DCNTTESIWA CLSCSHVACG
 51 RYIEEHALKH FQESSHPVAL EVNEMYVFCY LCDDYVLNDN ATGDLKLLRR
101 TLSAIKSQNY HCTTRSGRFL RSMGTGDDSY FLHDGAQSLL QSEDQLYTAL
151 WHRRRILMGK IFRTWFEQSP IGRKKQEEPF QEKIVVKREV KKRRQELEYQ
201 VKAELESMPP RKSLRLQGLA QSTIIEIVSV QVPAQTPASP AKDKVLSTSE
251 NEISQKVSDS SVKRRPIVTP GVTGLRNLGN TCYMNSVLQV LSHLLIFRQC
301 FLKLDLNQWL AMTASEKTRS CKHPPVTDTV VYQMNECQEK DTGFVCSRQS
351 SLSSGLSGGA SKGRKMELIQ PKEPTSQYIS LCHELHTLFQ VMWSGKWALV
401 SPFAMLHSVW RLIPAFRGYA QQDAQEFLCE LLDKIQRELE TTGTSLPALI
451 PTSQRKLIKQ VLNVVNNIFH GQLLSQVTCL ACDNKSNTIE PFWDLSLEFP
501 ERYQCSGKDI ASQPCLVTEM LAKFTETEAL EGKIYVCDQC NSKRRRFSSK
551 PVVLTEAQKQ LMICHLPQVL RLHLKRFRWS GRNNREKIGV HVGFEEILNM
601 EPYCCRETLK SLRPECFIYD LSAVVMHHGK GFGSGHYTAY CYNSEGGFWV
651 HCNDSKLSMC TMDEVCKAQA YILFYTQRVT ENGHSKLLPP ELLLGSQHPN
701 EDADTSSNEI LS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27dl, frame 2

PIR:S57591 hypothetical protein YMR223w - yeast (Saccharomyces cerevisiae), N = 4, Score = 218, P = 8.4e-38

SWISSPROT:UBPB_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13) (DEUBIQUITINATING ENZYME 11) (KIAA0055)., N = 2, Score = 300, P = 9.3e-31

TREMBL:AF079565_1 gene: "Ubp41"; product: "ubiquitin-specific protease UBP41"; Mus musculus ubiquitin-specific protease UBP41 (Ubp41) mRNA, complete cds., N = 3, Score = 187, P = 8.7e-30

PIR:I58376 hypothetical protein unp - mouse, N = 3, Score = 214, P = 1.2e-28

```
>SWISSPROT:UBPB_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15)
     (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13)
     (DEUBIQUITINATING ENZYME 11) (KIAA0055).
           Length = 1,118

  HSPs:

 Score = 300 (45.0 bits), Expect = 9.3e-31, Sum P(2) = 9.3e-31
 Identities = 95/301 (31%), Positives = 149/301 (49%)

Query:   381 LCHELHTLFQVMWSGKWALVSPFAMLHSVWRLIPAFRGYAQQDAQEFLCELLDKIQREL- 439
             +  E    + + +W+G++   +SP       ++ ++     F GY+QQD+QE L   L+D +   +L
Sbjct:   826 VAEEFGIIMKALWTGQYRYISPKDFKITIGKINDQFAGYSQQDSQELLLFLMDGLHEDLN 885

Query:   440 ---------ETTGTSLPALIPTSQRKLIKQVLN--VVNNIFHGQLLSQVTCLACDNKSNT 488
                      E      L                + LN  ++  +F GQ  S V CL C   KS T
Sbjct:   886 KADNRKRYKEENNDHLDDFKAAEHAWQKHKQLNESIIVALFQGQFKSTVQCLTCHKKSRT 945

Query:   489 IEPFWDLSLEFPERYQCSGKDIASQPCLVTEMLAKFTETEALEGKIYVCDQCNSKRRRFS 548
             E F  LSL         +C+ +D       CL    + +K   E +   +     + C   C  ++R
Sbjct:   946 FEAFMYLSLPLASTSKCTLQD-----CL--RLFSK--EEKLTDNNRFYCSHCRARR---- 992

Query:   549 SKPVVLTEAQKQLMICHLPQVLRLHLKRFRWSGRNNREKIGVHVGFE-EILNMEPYCC-- 605
                      ++ K++ I   LP VL +HLKRF + GR   ++K+     V F   E L++   Y
Sbjct:   993 -------DSLKKIEIWKLPPVLLVHLKRFSYDGRW-KQKLQTSVDFPLENLDLSQYVIGP 1044

Query:   606 RETLKSLRPECFIYDLSAVVMHHGKGFGSGHYTAYCYNSEGGFWVHCNDSKLSMCTMDEV 665
             +  LK         Y+L +V  H+G G    GHYTAYC N+     W    +D ++S  ++  V
Sbjct:  1045 KNNLKK-------YNLFSVSNHYG-GLDGGHYTAYCKNAARQRWFKFDDHEVSDISVSSV 1096

Query:   666 CKAQAYILFYTQ---RVTE 681
             + AYILFYT     RVT+
Sbjct:  1097 KSSAAYILFYTSLGPRVTD 1115

 Score = 126 (18.9 bits), Expect = 9.3e-31, Sum P(2) = 9.3e-31
 Identities = 41/116 (35%), Positives = 63/116 (54%)

Query:   200 QVKAELESMPPR--KSLRLQGLAQSTIIEIVSVQVPAQTPASPAKDKVLSTSENEISQKV 257
             Q+ AE +   P +  +S      + Q+  I+      + P  TP       ++K         + EIS ++
Sbjct:   701 QIPAERDREPSKLKRSYSSPDITQA--IQEEEKRKPTVTPTVNRENKPTCYPKAEIS-RL 757

Query:   258 SDSSVKR-RPIVT---PGVTGLRNLGNTCYMNSVLQVLS---HLLIF--RQCFLKLDLNQ 308
             S S ++   P+       P +TGLRNLGNTCYMNS+LQ L      HL   +   R C+     D+N+
Sbjct:   758 SASQIRNLNPVFGGSGPALTGLRNLGNTCYMNSILQCLCNAPHLADYFNRNCYQD-DINR 816

 Score = 50 (7.5 bits), Expect = 8.3e-23, Sum P(2) = 8.3e-23
 Identities = 29/106 (27%), Positives = 51/106 (48%)

Query:   173 RKKQEEPFQEKIVVKREVKKRRQELEYQVKAELESMPPRKSLRLQGLAQSTIIEIVSVQV 232
             + KQE+  +E+    +++ K R++E E + K + E+      +   Q  A+ + + S Q
Sbjct:   475 KNKQEKELRERQQEEQKEKLRKEEQEQKAKKKQEA-EENEITEKQQKAKEEMEKKESEQA 533

Query:   233 PAQ---TPASPAKD----KVLSTSENEIS--QKVSDSSVKRRPIVTPGV 272
             +     T A    K+      K   S SE+E S  +K   +    KR P  TP +
Sbjct:   534 KKEDKETSAKRGKEITGVKRQSKSEHETSDAKKSVEDRGKRCP--TPEI 580

 Score = 42 (6.3 bits), Expect = 5.7e-22, Sum P(2) = 5.7e-22
 Identities = 13/58 (22%), Positives = 27/58 (46%)

Query:   167 EQSPIGRKKQEEPFQEKIVVKREVKKRRQELEY-QVKAELESMPPRKSLRLQGLAQST 223
             EQ     +KKQE    E     +++ K+  ++ E  Q K E +     ++    + G+ + +
Sbjct:   498 EQEQKAKKKQEAEENEITEKQQKAKEEMEKKESEQAKKEDKETSAKRGKEITGVKRQS 555
```

Pedant information for DKFZphtes3_27d1, frame 2

Report for DKFZphtes3_27d1.2

[LENGTH] 712
[MW] 81155.71
[pI] 8.21
[HOMOL] SWISSPROT:UBPB_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13) (DEUBIQUITINATING ENZYME 11) (KIAA0055). 4e-32
[FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YMR223w] 5e-33
[FUNCAT] 06.07 protein modification (glycolsylation, acylation, myristylation, palmitylation, farnesylation and processing) [S. cerevisiae, YMR223w] 5e-33

```
[FUNCAT]      06.13 proteolysis     [S. cerevisiae, YBL067c] 3e-19
[FUNCAT]      10.03.99 other osmosensing activities      [S. cerevisiae, YDR069c] 2e-17
[FUNCAT]      03.10 sporulation and germination     [S. cerevisiae, YDR069c] 2e-17
[FUNCAT]      30.10 nuclear organization     [S. cerevisiae, YDR069c] 2e-17
[FUNCAT]      30.03 organization of cytoplasm      [S. cerevisiae, YDR069c] 2e-17
[FUNCAT]      09.25 vacuolar and lysosomal biogenesis      [S. cerevisiae, YDR069c] 2e-17
[FUNCAT]      04.05.01.04 transcriptional control   [S. cerevisiae, YNL186w] 4e-17
[FUNCAT]      99 unclassified proteins      [S. cerevisiae, YHL010c] 3e-12
[BLOCKS]      BL00970A Nuclear transition protein 2 proteins
[BLOCKS]      BL00972D
[BLOCKS]      BL00972C
[BLOCKS]      BL00972B
[BLOCKS]      BL00972A
[EC]          3.1.2.15 Ubiquitin thiolesterase 5e-06
[PIRKW]       alternative splicing 2e-11
[PIRKW]       thiolester hydrolase 5e-06
[PIRKW]       hydrolase 1e-14
[SUPFAM]      RING finger homology 7e-11
[SUPFAM]      deubiquinating enzyme SSV7 5e-16
[PROSITE]     MYRISTYL        5
[PROSITE]     AMIDATION       2
[PROSITE]     CAMP_PHOSPHO_SITE       1
[PROSITE]     CK2_PHOSPHO_SITE        10
[PROSITE]     TYR_PHOSPHO_SITE        2
[PROSITE]     UCH_2_2 1
[PROSITE]     PKC_PHOSPHO_SITE        17
[PROSITE]     ASN_GLYCOSYLATION       4
[PROSITE]     UCH_2_1 1
[PFAM]        Ubiquitin carboxyl-terminal hydrolases family 2
[PFAM]        Ubiquitin carboxyl-terminal hydrolases family 2
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY     4.92 %

SEQ     MLAMDTCKHVGQLQLAQDHSSLNPQKWHCVDCNTTESIWACLSCSHVACGRYIEEHALKH
SEG     ............................................................
PRD     cccccccchhhhhhhcccccccccceecccceeeeeeeecccccccccchhhhhhhhhhhh

SEQ     FQESSHPVALEVNEMYVFCYLCDDYVLNDNATGDLKLLRRTLSAIKSQNYHCTTRSGRFL
SEG     ............................................................
PRD     hhhhccceeecccceeeeeeccccccccccccchhhhhhhhhhhcccceeecccccccc

SEQ     RSMGTGDDSYFLHDGAQSLLQSEDQLYTALWHRRRILMGKIFRTWFEQSPIGRKKQEEPF
SEG     ............................................................
PRD     cccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhchhhhhhhhhh

SEQ     QEKIVVKREVKKRRQELEYQVKAELESMPPRKSLRLQGLAQSTIIEIVSVQVPAQTPASP
SEG     xxxxxxxxxxxxxxxx............................................
PRD     hheeehhhhhhhhhhhhhhhhhhhhccccccccccccccccceeeeeecccccccccccc

SEQ     AKDKVLSTSENEISQKVSDSSVKRRPIVTPGVTGLRNLGNTCYMNSVLQVLSHLLIFRQC
SEG     ............................................................
PRD     cccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhh

SEQ     FLKLDLNQWLAMTASEKTRSCKHPPVTDTVVYQMNECQEKDTGFVCSRQSSLSSGLSGGA
SEG     ..............................................xxxxxxxxxxxxxx
PRD     hhhhhchhhhhhhhhhhhhhhhccccccceehhhhccccccccccccccccccccccccc

SEQ     SKGRKMELIQPKEPTSQYISLCHELHTLFQVMWSGKWALVSPFAMLHSVWRLIPAFRGYA
SEG     xxxxx.......................................................
PRD     ccccceeeccccccchhhhhhhhhhhhhhhhhccceeeecccchhhhhhhhhhhhccccch

SEQ     QQDAQEFLCELLDKIQRELETTGTSLPALIPTSQRKLIKQVLNVVNNIFHGQLLSQVTCL
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhccccccccccchhhhhhhhhhhhhcccchhhhhhhhc

SEQ     ACDNKSNTIEPFWDLSLEFPERYQCSGKDIASQPCLVTEMLAKFTETEALEGKIYVCDQC
SEG     ............................................................
PRD     ccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhccceeecccc

SEQ     NSKRRRFSSKPVVLTEAQKQLMICHLPQVLRLHLKRFRWSGRNNREKIGVHVGFEEILNM
SEG     ............................................................
PRD     cccccccccccchhhhhhhhhhhhhhhchhhhhhhhhhhhhhccccccccccceeeecccccc

SEQ     EPYCCRETLKSLRPECFIYDLSAVVMHHGKGFGSGHYTAYCYNSEGGFWVHCNDSKLSMC
SEG     ............................................................
PRD     cccccccccccccccceeeeeeeeeeeeecccccccccceeeecccccccceeeeccccccc

SEQ     TMDEVCKAQAYILFYTQRVTENGHSKLLPPELLLGSQHPNEDADTSSNEILS
SEG     ....................................................
PRD     cchhhhhhhhhhhhhhheeeeccccccccccccccccccccccccccccccc
```

Prosite for DKFZphtes3_27d1.2

PS00001	33->37	ASN_GLYCOSYLATION	PDOC00001
PS00001	90->94	ASN_GLYCOSYLATION	PDOC00001
PS00001	484->488	ASN_GLYCOSYLATION	PDOC00001
PS00001	653->657	ASN_GLYCOSYLATION	PDOC00001
PS00004	545->549	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	113->116	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005
PS00005	315->318	PKC_PHOSPHO_SITE	PDOC00005
PS00005	320->323	PKC_PHOSPHO_SITE	PDOC00005
PS00005	394->397	PKC_PHOSPHO_SITE	PDOC00005
PS00005	453->456	PKC_PHOSPHO_SITE	PDOC00005
PS00005	506->509	PKC_PHOSPHO_SITE	PDOC00005
PS00005	542->545	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	580->583	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	611->614	PKC_PHOSPHO_SITE	PDOC00005
PS00005	676->679	PKC_PHOSPHO_SITE	PDOC00005
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	223->227	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	249->253	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	525->529	CK2_PHOSPHO_SITE	PDOC00006
PS00006	661->665	CK2_PHOSPHO_SITE	PDOC00006
PS00006	706->710	CK2_PHOSPHO_SITE	PDOC00006
PS00007	193->200	TYR_PHOSPHO_SITE	PDOC00007
PS00007	192->200	TYR_PHOSPHO_SITE	PDOC00007
PS00008	218->224	MYRISTYL	PDOC00008
PS00008	355->361	MYRISTYL	PDOC00008
PS00008	359->365	MYRISTYL	PDOC00008
PS00008	471->477	MYRISTYL	PDOC00008
PS00008	589->595	MYRISTYL	PDOC00008
PS00009	171->175	AMIDATION	PDOC00009
PS00009	362->366	AMIDATION	PDOC00009
PS00972	274->290	UCH_2_1	PDOC00750
PS00973	619->638	UCH_2_2	PDOC00750

Pfam for DKFZphtes3_27d1.2

```
HMM_NAME    Ubiquitin carboxyl-terminal hydrolases family 2

HMM             *GIqNlGNTCYMNSIIQCL*
                 G++NLGNTCYMNS++Q+L
Query       274  GLRNLGNTCYMNSVLQVL    291


HMM_NAME    Ubiquitin carboxyl-terminal hydrolases family 2

HMM             *YdLYgVICHYGntldyGHYWaYVKNenhHRWkWYYFDDEtV*
                 YDL +V+ H+G + ++GHY+AY++N +    ++W+  +D++
Query       619  YDLSAVVMHHGKGFGSGHYTAYCYNSE--GGFWVHCNDSKL    657
```

DKFZphtes3_27k4
----

group: transmembrane protein

Summary DKFZphtes3_27k4 encodes a novel 490 amino acid protein with similarity to two hypothetical C.elegans proteins.

The novel protein contains 10 transmembrane regions and a leucine zipper. It is a member of the new 10 trans-membrane domain containing protein family which is specific for multicellular eukariotes.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

strong similarity to C.elegans K07H8.2/ZK185.2
membrane regions: 10

complete cDNA, complete cds potential start at Bp 109, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1901 bp
Poly A stretch at pos. 1866, no polyadenylation signal found

```
   1 GTGATTTACC AGAAAAACCA AGAAGACAGG CACAAAAAAG CAAACGGCAT
  51 TTGGCAAGAT GGATTATCAA CTGCAGTACA GACTTTTAGT AATAGATCTG
 101 AGCAACACAT GGAGTATCAC AGTTTCTCAG AGCAGTCTTT TCATGCCAAT
 151 AATGGGCACG CATCATCAAG CTGCAGCCAA AAGTATGATG ACTATGCCAA
 201 TTATAATTAC TGTGATGGAA GGGAGACTTC AGAAACCACT GCCATGTTAC
 251 AAGATGAAGA TATATCTAGT GATGGTGATG AAGATGCTAT TGTAGAAGTG
 301 ACCCCAAAAT TACCAAAGGA ATCCAGTGGC ATCATGGCAT TGCAAATACT
 351 TGTGCCCTTT TTGCTAGCTG GTTTTGGAAC AGTTTCAGCT GGCATGGTAC
 401 TGGATATAGT ACAGCACTGG GAGGTGTTCA GAAAAGTTAC AGAAGTTTTC
 451 ATTTTAGTCC CTGCACTTCT TGGTCTCAAA GGGAACTTGG AAATGACATT
 501 GGCATCCAGA TTATCCACTG CAGTAAATAT TGGGAAGATG GATTCACCCA
 551 TTGAAAAGTG GAACCTAATA ATTGGCAACT TGGCTTTAAA GCAGGTTCAG
 601 GCAACAGTAG TGGGTTTTCT AGCAGCTGTG GCAGCAATTA TATTGGGCTG
 651 GATTCCAGAA GGAAAATATT ACCTTGATCA TTCCATACTT CTGTGCTCTA
 701 GCAGTGTGGC AACTGCCTTC ATTGCATCTC TTCTGCAGGG AATAATAATG
 751 GTTGGGGTTA TCGTTGGTTC AAAGAAGACT GGTATAAATC CTGATAATGT
 801 TGCTACACCC ATTGCTGCTA GTTTTGGCGA CCTTATAACT CTTGCCATAT
 851 TGGCTTGGAT AAGTCAGGGC TTATACTCCT GTCTTGAGAC CTATTACTAC
 901 ATTTCTCCAT TAGTTGGTGT ATTTTTCTTG GCTCTAACCC CTATTTGGAT
 951 TATAATAGCT GCCAAACATC CAGCCACAAG AACAGTTCTC CACTCAGGCT
1001 GGGAGCCTGT CATAACAGCT ATGGTTATAA GTAGCATTGG GGGCCTTATT
1051 CTGGACACAA CTGTATCAGA CCCAAACTTG GTTGGGATTG TTGTTTACAC
1101 GCCAGTTATT AATGGTATTG GTGGTAATTT GGTGGCCATT CAGGCTAGCA
1151 GGATTTCTAC CTACCTCCAT TTACATAGCA TTCCAGGAGA ATTGCCTGAT
1201 GAACCCAAAG GTTGTTACTA CCCATTTAGA ACTTTCTTTG GTCCAGGAGT
1251 AAATAATAAG TCTGCTCAAG TTCTACTGCT TTTAGTGATT CCTGGACATT
1301 TAATTTTCCT CTACACTATT CATTTGATGA AAAGTGGTCA TACTTCTTTA
1351 ACTATAATCT TCATAGTAGT GTATTTATTT GGCGCTGTGT TACAGGTATT
1401 TACCTTGCTG TGGATTGCTG ACTGGATGGT CCATCACTTC TGGAGGAAAG
1451 GAAAGGACCC GGATAGTTTC TCCATCCCCT ACCTAACAGC ATTGGGTGAT
1501 CTGCTCGGGA CAGCTCTGTT AGCCTTAAGT TTTCATTTTC TTTGGCTTAT
1551 TGGAGATCGA GATGGAGATG TTGGAGACTA ATAAATTCTA CAAACTGCTC
1601 TCAAGTTACC AAGGAAGAAA ATACACGACA ACCACTTATG GCTCTTTTTC
1651 AAAACTCTTA AATCAGTAGT TTGACTTTTG CCAGGGTAAT CTTCAGTTGG
1701 CCCTGATTCA ATTAAATGGC CTTAATTTTT TTTTAAGGAA TTTGTGTCAA
1751 AACCAGAATG AAGAGTATTC GTGCTGCTTT TCATAGAATA AATGATAATT
1801 TGACATAGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1851 AAAAAAAAAA AAGGGGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAGGG
1901 G
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 109 bp to 1578 bp; peptide length: 490
Category: similarity to unknown protein

```
  1 MEYHSFSEQS FHANNGHASS SCSQKYDDYA NYNYCDGRET SETTAMLQDE
 51 DISSDGDEDA IVEVTPKLPK ESSGIMALQI LVPFLLAGFG TVSAGMVLDI
101 VQHWEVFRKV TEVFILVPAL LGLKGNLEMT LASRLSTAVN IGKMDSPIEK
151 WNLIIGNLAL KQVQATVVGF LAAVAAIILG WIPEGKYYLD HSILLCSSSV
201 ATAFIASLLQ GIIMVGVIVG SKKTGINPDN VATPIAASFG DLITLAILAW
251 ISQGLYSCLE TYYYISPLVG VFFLALTPIW IIIAAKHPAT RTVLHSGWEP
301 VITAMVISSI GGLILDTTVS DPNLVGIVVY TPVINGIGGN LVAIQASRIS
351 TYLHLHSIPG ELPDEPKGCY YPFRTFFGPG VNNKSAQVLL LLVIPGHLIF
401 LYTIHLMKSG HTSLTIIFIV VYLFGAVLQV FTLLWIADWM VHHFWRKGKD
451 PDSFSIPYLT ALGDLLGTAL LALSFHFLWL IGDRDGDVGD
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27k4, frame 1

TREMBL:AF036704_2 gene: "ZK185.2"; Caenorhabditis elegans cosmid ZK185., N = 1, Score = 730, P = 3.1e-72

TREMBL:AF047659_9 gene: "K07H8.2"; Caenorhabditis elegans cosmid K07H8., N = 1, Score = 940, P = 1.7e-94

>TREMBL:AF047659_9 gene: "K07H8.2"; Caenorhabditis elegans cosmid K07H8.
Length = 507

HSPs:

Score = 940 (141.0 bits), Expect = 1.7e-94, P = 1.7e-94
Identities = 204/412 (49%), Positives = 271/412 (65%)

```
Query:   68 LPKESSGIMALQILVPFLLAGFGTVSAGMVLDIVQHWEVFRKVTEVFILVPALLGLKGNL 127
            +P ESS ++  Q+L PF +AG G V AG+VL IV  W +F ++ E+ ILVPALLGLKGNL
Sbjct:   82 IPAESSYVLFFQVLFPFAVAGLGMVFAGLVLSIVVTWPLFEEIPEILILVPALLGLKGNL 141

Query:  128 EMTLASRLSTAVNIGKMDSPIEKWNLIIGNLALKQVQATVVGFLAAVAAIILGWIPEGKY 187
            EMTLASRLST  N+G MDS  ++ +++I NLAL QVQATVV FLA+  A  L +IP G +
Sbjct:  142 EMTLASRLSTLANLGHMDSSKQRKDVVIANLALVQVQATVVAFLASAFAAALAFIPSGDF 201

Query:  188 YLDHSILLCSSSVATAFIASLLQGIIMVGVIVGSKKTGINPDNVATPIAASFGDLITLAI 247
              H L+C+SS+ATA  ASL+  ++MV VIV S+K  INPDNVATPIAAS GDL TL +
Sbjct:  202 DWAHGALMCASSLATACSASLVLSLLMVVVIVTSRKYNINPDNVATPIAASLGDLTTLTV 261

Query:  248 LAWISQGLYSCLETYYYISPLVGVFFLALTPIWIIIAAKHPATRTVLHSGWEPVITAMVI 307
            LA+           T  +++ +V V FL L P WI IA ++  T+  L++GW PVI +M+I
Sbjct:  262 LAFFGSVFLKAHNTESWLNVIVIVLFLLLLPFWIKIANENEGTQETLYNGWTPVIMSMLI 321

Query:  308 SSIGGLILDTTVSDPNLVGIVVYTPVINGIGGNLVAIQASRISTYLHLHSIPGELPDEPK 367
            SS GG IL+T V    +   + Y PV+NG+GGNL A+QASR+STY H      G LP+E
Sbjct:  322 SSAGGFILETAVRRYH--SLSTYGPVLNGVGGNLAAVQASRLSTYFHKAGTVGVLPNEWT 379

Query:  368 GCYYPF--RTFFGPGVNNKSAQVLLLLVIPGHLIFLYTIHLM----KSGHTSLTIIFIVV 421
              +     R FF    +++SA+VLLLLV+PGH+ F + I L      K+  T     +F  +
Sbjct:  380 VSRFTSVQRAFFSKEWDSRSARVLLLLVVPGHICFNFLIQLFTLTSKNNVTPHGPLFTSL 439

Query:  422 YLFGAVLQVFTLLWIADWMVHHFWRKGKDPDSFSIPYLTALGDLLGTALLALSF 475
            Y+  A++QV  LL++    +V   W+   DPD+  IPYLTALGDLLGT LL + F
Sbjct:  440 YMIAAIIQVVILLFVCQLLVALLWKWKIDPDNSVIPYLTALGDLLGTGLLFIVF 493
```

Pedant information for DKFZphtes3_27k4, frame 1

Report for DKFZphtes3_27k4.1

[LENGTH] 490
[MW] 53266.39

[pI]	5.29	
[HOMOL]	TREMBL:AF047659_9 gene: "K07H8.2"; Caenorhabditis elegans cosmid K07H8. 4e-94	
[PROSITE]	LEUCINE_ZIPPER	1
[PROSITE]	MYRISTYL	7
[PROSITE]	CAMP_PHOSPHO_SITE	1
[PROSITE]	CK2_PHOSPHO_SITE	7
[PROSITE]	PROKAR_LIPOPROTEIN	2
[PROSITE]	TYR_PHOSPHO_SITE	1
[PROSITE]	PKC_PHOSPHO_SITE	3
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	TRANSMEMBRANE	10
[KW]	LOW_COMPLEXITY	3.06 %

```
SEQ  MEYHSFSEQSFHANNGHASSSCSQKYDDYANYNYCDGRETSETTAMLQDEDISSDGDEDA
SEG  ............................................................
PRD  ccccccceeeccccccccccccccccccccceeeccccccccchhhhhhhcccccccccee
MEM  ............................................................

SEQ  IVEVTPKLPKESSGIMALQILVPFLLAGFGTVSAGMVLDIVQHWEVFRKVTEVFILVPAL
SEG  ............................................................
PRD  eeeecccccchhhhhhhhhhhhhhhhccchhhhhhhhhcchhhhhcccceeeeeeeeccc
MEM  ................MMMMMMMMMMMMMMMMMMMMMMMMMMMMM.MMMMMMMMMMMMMM

SEQ  LGLKGNLEMTLASRLSTAVNIGKMDSPIEKWNLIIGNLALKQVQATVVGFLAAVAAIILG
SEG  ............................................................
PRD  ccccchhhhhhhhhhhhhccccccccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  MMMMMMM.........................MMMMMMMMMMMMMMMMMMMM........

SEQ  WIPEGKYYLDHSILLCSSSVATAFIASLLQGIIMVGVIVGSKKTGINPDNVATPIAASFG
SEG  ............................................................
PRD  hcccceeecccceeehhhhhhhhhhhhhhhhhhhhheeeeccccccccccccccccccccc
MEM  ............MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM............MMMMMM

SEQ  DLITLAILAWISQGLYSCLETYYYISPLVGVFFLALTPIWIIIAAKHPATRTVLHSGWEP
SEG  ............................................................
PRD  cchhhhhhhhhhhhhhhhcceeeeehhhhhhhhhhchhhhhhhhccccccccchhhhhhh
MEM  MMMMMMMMMMMMMMMM....MMMMMMMMMMMMMMMMMMMMM.............MMMMMM

SEQ  VITAMVISSIGGLILDTTVSDPNLVGIVVYTPVINGIGGNLVAIQASRISTYLHLHSIPG
SEG  ............................................................
PRD  hcchhhhhcceeeeccccccccceeeeeeceeeeccccccceeeeehhhhhhhhhhcccc
MEM  MMMMMMMMMMMMMMMMM...........................................

SEQ  ELPDEPKGCYYPFRTFFGPGVNNKSAQVLLLLVIPGHLIFLYTIHLMKSGHTSLTIIFIV
SEG  ............................................................
PRD  cccccccccccceeeeeccccchhhhhhhhhhccccchhhhhhhhccccccccceeehhh
MEM  .....................MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...MMMMMM

SEQ  VYLFGAVLQVFTLLWIADWMVHHFWRKGKDPDSFSIPYLTALGDLLGTALLALSFHFLWL
SEG  .....................................xxxxxxxxxxxxxxxx.......
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccceeeeeecchhhhhhhhhhhheeee
MEM  MMMMMMMMMMMMMMMMMMMMMM..............MMMMMMMMMMMMMMMMMMMMMMMM

SEQ  IGDRDGDVGD
SEG  ..........
PRD  eecccccccc
MEM  MM........
```

Prosite for DKFZphtes3_27k4.1

PS00001	383->387	ASN_GLYCOSYLATION	PDOC00001
PS00004	108->112	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	23->26	PKC_PHOSPHO_SITE	PDOC00005
PS00005	65->68	PKC_PHOSPHO_SITE	PDOC00005
PS00005	221->224	PKC_PHOSPHO_SITE	PDOC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	296->300	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00007	25->33	TYR_PHOSPHO_SITE	PDOC00007
PS00008	90->96	MYRISTYL	PDOC00008
PS00008	122->128	MYRISTYL	PDOC00008
PS00008	216->222	MYRISTYL	PDOC00008
PS00008	220->226	MYRISTYL	PDOC00008

PS00008	254->260	MYRISTYL	PDOC00008
PS00008	336->342	MYRISTYL	PDOC00008
PS00008	339->345	MYRISTYL	PDOC00008
PS00013	12->23	PROKAR_LIPOPROTEIN	PDOC00013
PS00013	248->259	PROKAR_LIPOPROTEIN	PDOC00013
PS00029	459->481	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_27k4.1)

DKFZphtes3_27o14
----

group: testes derived

DKFZphtes3_27o14 encodes a novel 358 amino acid protein with similarity to C. elegans cosmid C55A6.

The new protein contains a C3HC4 zinc finger (RING finger) signature. The ring finger structure binds two atoms of zinc, and is involved in mediating protein-protein interactions. No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans C55A6.1

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: /map="6"

Insert length: 2158 bp
Poly A stretch at pos. 2137, polyadenylation signal at pos. 2120

```
   1 CCGAGGCCAG AGAGAAAAGA CTGCGAGGTG GCCGCAGCTG TGGCCGGAGA
  51 GCACAAAGAA TGAACCAGCA GTGGAAGAGA AAATACTGTA AGCTGGCTGA
 101 CTGCTGGTGA AGAAAATGCT TTATTTTTGT GGCAGGCATC TGTGGGATCT
 151 GTAATAGAAA TATATTGGAG TAATTCAAGA TTCTGTGGTT GGCCCTTTTG
 201 ACTGCTCTCT CTACAGGTTT AATTTGGGCA TTTACTCATT TTCATGGCTC
 251 CAAGGACCAT GTATGTGTTG GGGATCTTCA ATATTCATGT TATTTTCTCC
 301 TTTGGTCTTA TATGATTGTT ACCTTTATGA AGCTTTAGTG ATTACAAAGC
 351 ACTTTTTTTG TCCATTTTTA CCTGAGCTTT GTAAACTCTG ATTTGCAGGA
 401 TGGCTGGCTG TGGTGAAATT GATCATTCAA TAAACATGCT TCCTACAAAC
 451 AGGAAAGCGA ACGAGTCCTG TTCTAATACT GCACCTTCTT TAACCGTCCC
 501 TGAATGTGCC ATTTGTCTGC AAACATGTGT TCATCCAGTC AGTCTGCCCT
 551 GTAAGCACGT TTTCTGCTAT CTATGTGTAA AAGGAGCTTC ATGGCTTGGA
 601 AAGCGGTGTG CTCTTTGTCG ACAAGAAATT CCCGAGGATT TCCTTGACAA
 651 GCCAACCTTG TTGTCACCAG AAGAACTCAA GGCAGCAAGT AGAGGAAATG
 701 GTGAATATGC ATGGTATTAT GAAGGAAGAA ATGGGTGGTG GCAGTACGAT
 751 GAGCGCACTA GTAGAGAGCT GGAAGATGCT TTTTCCAAAG GTAAAAAGAA
 801 CACTGAAATG TTAATTGCTG GCTTTCTGTA TGTCGCTGAT CTTGAAAACA
 851 TGGTTCAATA TAGGAGAAAT GAACATGGAC GTCGCAGGAA GATTAAGCGA
 901 GATATAATAG ATATACCAAA GAAGGGAGTA GCTGGACTTA GGCTAGACTG
 951 TGATGCTAAT ACCGTAAACC TAGCAAGAGA GAGCTCTGCT GACGGAGCGG
1001 ACAGTGTATC AGCACAGAGT GGAGCTTCTG TTCAGCCCCT AGTGTCTTCT
1051 GTAAGGCCCC TAACATCAGT AGATGGTCAG TTAACAAGCC CTGCAACACC
1101 ATCCCCTGAT GCAAGCACTT CTCTGGAAGA CTCTTTTGCT CATTTACAAC
1151 TCAGTGGAGA CAACACAGCT GAAAGGAGTC ATAGGGGAGA AGGAGAAGAA
1201 GATCATGAAT CACCATCTTC AGGCAGGGTA CCAGCACCAG ACACCTCCAT
1251 TGAAGAAACT GAATCAGATG CCAGTAGTGA TAGTGAGGAT GTATCTGCAG
1301 TTGTTGCACA GCACTCCTTG ACCCAACAGA GACTTTTGGT TTCTAATGCA
1351 AACCAGACAG TACCCGATCG ATCAGATCGA TCGGGAACTG ATCGATCAGT
1401 AGCAGGGGGT GGAACAGTGA GTGTCAGTGT CAGATCTAGA AGGCCTGATG
1451 GACAGTGCAC AGTAACTGAA GTTTAAATAA AAATGTCTTC AGCTCCATGC
1501 TCAAGGTTGA AAGGGTTACC TGTAAATTTC TGCCCACATA ACATTATACT
1551 CATCCCTAGT AGTGCATTTT GGGAGTTGGG GTGGGAAGGG GTATGGGAAG
1601 GATAGACTCA TAATTAAAAT GTCTAACATG TCTCTGTTGA GAAATTTATT
1651 TAATGTAAGG AACTTGGGTG TTAATAGTTG AGAGCTGTTT AGTAATAACC
1701 CAGTTTTCTT GAGGTCTGTT TACTTTATAC TTTTTAAAAA CTTCTGTAGT
1751 TCTTTTGGCC AGTGTGTTTG TATTATCTGT GCATTAATGG TCCTCATCTG
1801 ACTCCTGCAT TGTGTCTTAT TTTTCTGCAT GGATTGGCAT AAGACCATTA
1851 CTAAAATTTG GCACCTGTGA GATGTTTGAT ATTATGAACA GGAAACATAA
1901 TTTAATGTAT GAATAGATGT GAATTTGGGA TTTCAAAATA GATGAATAAC
1951 AACTATTTTA TAGTAAAGTT ATTGAAATGG AAATGAAAAC AGCCAGTAAC
2001 TTATGTTTCA GAATGTTTGT AACACACTTC ATGGTGTTCC CATAGGCTTT
2051 GCTGTCTAGT CTTATAGTTT GAGGTTTTTT TGGTCTGCAT TTTTCTTTTT
2101 GATTACAAAA TTTATAATTT AATAAATACT AGAGTTTATC AAAAAAAAAA
2151 AAAAAAAG
```

BLAST Results

Entry HSG117 from database EMBL:
human STS SHGC-36270.
Score = 1148, P = 8.9e-45, identities = 240/250

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 400 bp to 1473 bp; peptide length: 358
Category: similarity to unknown protein
Prosite motifs: ZINC_FINGER_C3HC4 (51-61)

```
  1 MAGCGEIDHS INMLPTNRKA NESCSNTAPS LTVPECAICL QTCVHPVSLP
 51 CKHVFCYLCV KGASWLGKRC ALCRQEIPED FLDKPTLLSP EELKAASRGN
101 GEYAWYYEGR NGWWQYDERT SRELEDAFSK GKKNTEMLIA GFLYVADLEN
151 MVQYRRNEHG RRRKIKRDII DIPKKGVAGL RLDCDANTVN LARESSADGA
201 DSVSAQSGAS VQPLVSSVRP LTSVDGQLTS PATPSPDAST SLEDSFAHLQ
251 LSGDNTAERS HRGEGEEDHE SPSSGRVPAP DTSIEETESD ASSDSEDVSA
301 VVAQHSLTQQ RLLVSNANQT VPDRSDRSGT DRSVAGGGTV SVSVRSRRPD
351 GQCTVTEV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_27o14, frame 1

TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6,
N = 2, Score = 165, P = 4.2e-15

SWISSPROT:YWZ6_CAEEL HYPOTHETICAL 39.3 KD PROTEIN C02B8.6 IN CHROMOSOME
X., N = 2, Score = 136, P = 3.1e-11

```
>TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6
          Length = 484

  HSPs:

 Score = 165 (24.8 bits), Expect = 4.2e-15, Sum P(2) = 4.2e-15
 Identities = 42/106 (39%), Positives = 61/106 (57%)

Query:    75 QEIPEDFLDKPTLLSPEELKAASRGNGEYAWYYEGRN-GWWQYDERTSRELEDAFSKGKK 133
             Q +P   LD  ++  PEE K         Y W Y G+N GWW+++ R  RE+E+A++ GK
Sbjct:    93 QNVPALDLDA-SICDPEERK--------Y-WIYSGKNQGWWRFEPRNEREIEEAYNAGKC 142

Query:   134 NTEMLIAGFLYVADLENMVQYRRNEHGRRRKIKR---DIID-IPKKGVAGL 180
             + E++I G  YV D    +QY R    + R +KR   D  D I  KG+AG+
Sbjct:   143 HCEVVICGRPYVIDFHQFLQYPRGVPNQARHVKRVSADDFDGIGVKGLAGI 193

 Score = 96 (14.4 bits), Expect = 4.2e-15, Sum P(2) = 4.2e-15
 Identities = 19/54 (35%), Positives = 30/54 (55%)

Query:    35 ECAICLQTCVHPVSLP-CKHVFCYLCVKGASW--LGKRCALCRQEIPEDFLDKPT 86
             EC IC     + P ++P C H FC++C+KG     +G  C +CR  I  +   +P+
Sbjct:    11 ECPICQCKMIVPTTIPACGHKFCFICLKGVYMNDMGG-CPMCRGPIDSNIFAQPS 64
```

Pedant information for DKFZphtes3_27o14, frame 1

Report for DKFZphtes3_27o14.1

```
[LENGTH]        358
[MW]            38818.90
[pI]            5.17
[HOMOL]         TREMBL:CEC55A6_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6 2e-12

[FUNCAT]        11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair)         [S. cerevisiae, YCR066w] 3e-04
[FUNCAT]        03.19 recombination and dna repair    [S. cerevisiae, YCR066w] 3e-04
[FUNCAT]        30.10 nuclear organization     [S. cerevisiae, YCR066w] 3e-04
```

```
[FUNCAT]        06.07 protein modification (glycolsylation, acylation, myristylation,
palmitylation, farnesylation and processing)        [S. cerevisiae, YCR066w] 3e-04
[FUNCAT]        06.10 assembly of protein complexes  [S. cerevisiae, YDR265w] 4e-04
[FUNCAT]        30.19 peroxisomal organization       [S. cerevisiae, YDR265w] 4e-04
[BLOCKS]        BL00518 Zinc finger, C3HC4 type, proteins
[PROSITE]       MYRISTYL        2
[PROSITE]       AMIDATION       3
[PROSITE]       CAMP_PHOSPHO_SITE       1
[PROSITE]       CK2_PHOSPHO_SITE        12
[PROSITE]       TYR_PHOSPHO_SITE        1
[PROSITE]       ZINC_FINGER_C3HC4       1
[PROSITE]       PKC_PHOSPHO_SITE        9
[PROSITE]       ASN_GLYCOSYLATION       2
[PFAM]          Zinc finger, C3HC4 type (RING finger)
[KW]            Irregular
[KW]            3D
[KW]            LOW_COMPLEXITY    19.83 %

SEQ     MAGCGEIDHSINMLPTNRKANESCSNTAPSLTVPECAICLQTCVHPVSLPCKHVFCYLCV
SEG     ............................................................
1rmd-   ................................TTTTEETTTEEEETTTEEEEHHHH

SEQ     KGASWLGKRCALCRQEIPEDFLDKPTLLSPEELKAASRGNGEYAWYYEGRNGWWQYDERT
SEG     ............................................................
1rmd-   HHHHHHCCBTTTTTCBCGGG-CBCC...................................

SEQ     SRELEDAFSKGKKNTEMLIAGFLYVADLENMVQYRRNEHGRRRKIKRDIIDIPKKGVAGL
SEG     ....................................xxxxxxxxxxxxxxx.....
1rmd-   ............................................................

SEQ     RLDCDANTVNLARESSADGADSVSAQSGASVQPLVSSVRPLTSVDGQLTSPATPSPDAST
SEG     ..............................................xxxxxxxxxxxx
1rmd-   ............................................................

SEQ     SLEDSFAHLQLSGDNTAERSHRGEGEEDHESPSSGRVPAPDTSIEETESDASSDSEDVSA
SEG     x...................................xxxxxxxxxxxxxxxxxxxx
1rmd-   ............................................................

SEQ     VVAQHSLTQQRLLVSNANQTVPDRSDRSGTDRSVAGGGTVSVSVRSRRPDGQCTVTEV
SEG     xxx...........................xxxxxxxxxxxxxxxxxxx.......
1rmd-   ..........................................................
```

Prosite for DKFZphtes3_27o14.1

PS00001	21->25	ASN_GLYCOSYLATION	PDOC00001
PS00001	318->322	ASN_GLYCOSYLATION	PDOC00001
PS00004	132->136	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	120->123	PKC_PHOSPHO_SITE	PDOC00005
PS00005	217->220	PKC_PHOSPHO_SITE	PDOC00005
PS00005	260->263	PKC_PHOSPHO_SITE	PDOC00005
PS00005	274->277	PKC_PHOSPHO_SITE	PDOC00005
PS00005	325->328	PKC_PHOSPHO_SITE	PDOC00005
PS00005	330->333	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	120->124	CK2_PHOSPHO_SITE	PDOC00006
PS00006	195->199	CK2_PHOSPHO_SITE	PDOC00006
PS00006	222->226	CK2_PHOSPHO_SITE	PDOC00006
PS00006	240->244	CK2_PHOSPHO_SITE	PDOC00006
PS00006	282->286	CK2_PHOSPHO_SITE	PDOC00006
PS00006	287->291	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00007	98->107	TYR_PHOSPHO_SITE	PDOC00007
PS00008	329->335	MYRISTYL	PDOC00008
PS00008	337->343	MYRISTYL	PDOC00008
PS00009	66->70	AMIDATION	PDOC00009
PS00009	130->134	AMIDATION	PDOC00009
PS00009	159->163	AMIDATION	PDOC00009
PS00518	51->61	ZINC_FINGER_C3HC4	PDOC00449

Pfam for DKFZphtes3_27o14.1

```
HMM_NAME        Zinc finger, C3HC4 type (RING finger)

HMM                 *CPICFcTFQlDyPWPFdePmMlPCgHsFCypCIrrW......CPmC*
                     C+IC      L      + P++LPC+H+FCY C++         C +C
Query            36  CAIC-----LQT---CVHPVSLPCKHVFCYLCVKGASWLGKRCALC      73
```

DKFZphtes3_28d14

group: testes derived

DKFZphtes3_28d14 encodes a novel 97 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1279 bp
Poly A stretch at pos. 1232, no polyadenylation signal found

```
   1 GGAGCTCAGA AGTTGGGCAA AGGTCACAGC AGACTTCCTG AAAAGCAGAC
  51 ACTGAGGAAC ACAGTGGAGA GCGGGAGTTC ACAGCGACGC AGCTGAGGAC
 101 GACGCAGGAC CTCTCCCAAA GGTGCTGCAG CTCCAGCACC AGGGGCCAGG
 151 GCTGCGGCGA CAGCAGCTCA GCAACCCTTG CTGTGCTCAA GTTCTTGGGG
 201 ATTCAGAGCT AAGTTCAAAA TTTAGAAACA GTGCCTTAAA GACGGGCAAG
 251 AAAACCCGGT GTGGGAGTCT GCTCATCTAT GGTTTGTTAC TGCTCTCGCT
 301 TTGATATTCT TAAATTCCTA GGTACCAATG AAAAAGCCAA GTGAACGTGG
 351 CAGAGTGAGG AGGAGACAGG AGCGTGTGCA CCTTCCATCT GTGAGAGGCA
 401 CACTTCAGTC TGGGTTCAAG ATGCAGAATG GTGCCTACAG CAAAAAAAAA
 451 AAAAACACCC TCCTCCCTTC TTTACCATTT GAATGGACAT TTTCCTTACC
 501 TGTGATCCCA ACAGAAACAG ATCCAGACCT ATCATGTGAA GTCCACGTTC
 551 CAGGATCAGA AGTAACCAGT TTATGGACTG AGCTTACACG GGAAAGTCTA
 601 CCCCCGACTC CTTCTGGATA GTAACATACA CAGCTGCATA AAAACGTCTC
 651 CAAGGGGACA TACGATGCAT TTGCTTGGTG TCCCAGCCAA GCTCCCCACC
 701 GGCGACCTCA CTGTTCCTTA GAGCTCGAGA GCTCGTCTCC TATCAATCAG
 751 AGAACCCCAT CAGCTGTGAC CAACAGAGCT GGAGCCCTCT GTGGAGGGAG
 801 CTGACCCCAC ACACAGGACA GAGCAGAATC CTGATTATTT TACAAACTGC
 851 AAACCTTCTG AGTAAGAAGA CAAAAATATA CATTCCAAGG TATCTGTAAA
 901 GTGCTTGGAA GATGCAGACA GCTGCACCGA GGGGCTCTGA TCCATCCACA
 951 CGCTGCGCTT TGCTGCGGTC ACACACACGG TCTCAGTCAC GTGATGGTTT
1001 TGCTTTTATT TCTTAAACGG CTGAGTGATA ATCCAGCTAG TGTGCAGTCA
1051 TTTCATACCT TTCAATGGGC GTCACCGCAG TGACGCTGCC CCAGCCCCAT
1101 GCTGAGGGCC GACACAATTC ACGGAACAGA TTCATCATAT TTGGTCTTTA
1151 TGTAAATAAT AAATGTTTTA AAATTGCCTA AATATAAAAA AAAAAAAAAA
1201 AAAAAAAAAA AAAAAAAAAA AAAGGGCGGC CGAAAAAAAA AAAAAAAAAA
1251 AAAAAAAAAA AAAAAAAAAA GGGCGGCCG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 328 bp to 618 bp; peptide length: 97
Category: putative protein

```
 1 MKKPSERGRV RRRQERVHLP SVRGTLQSGF KMQNGAYSKK KKNTLLPSLP
51 FEWTFSLPVI PTETDPDLSC EVHVPGSEVT SLWTELTRES LPPTPSG
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_28d14, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_28d14, frame 1

Report for DKFZphtes3_28d14.1

[LENGTH]	97	
[MW]	10945.56	
[pI]	9.80	
[PROSITE]	MYRISTYL	2
[PROSITE]	CAMP_PHOSPHO_SITE	2
[PROSITE]	CK2_PHOSPHO_SITE	2
[PROSITE]	PKC_PHOSPHO_SITE	3
[KW]	All_Alpha	
[KW]	LOW_COMPLEXITY	12.37 %

```
SEQ    MKKPSERGRVRRRQERVHLPSVRGTLQSGFKMQNGAYSKKKKNTLLPSLPFEWTFSLPVI
SEG    .....xxxxxxxxxxxx...........................................
PRD    cccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccc

SEQ    PTETDPDLSCEVHVPGSEVTSLWTELTRESLPPTPSG
SEG    .....................................
PRD    ccccccccceeeeccccccchhhhhhhhhhcccccccc
```

Prosite for DKFZphtes3_28d14.1

PS00004	2->6	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	41->45	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00006	62->66	CK2_PHOSPHO_SITE	PDOC00006
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	76->82	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_28d14.1)

DKFZphtes3_2a11
----

group: testes derived

DKFZphtes3_2a11 encodes a novel 1048 amino acid protein with very weak similarity to mucins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to mucin

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 4082 bp
Poly A stretch at pos. 4060, polyadenylation signal at pos. 4034

```
   1 GAGGACTGCG AGCACAGCGG CGGCCGGGTG GCGGGGGTGA GTGGGGCCAG
  51 CGGGGCTGGA CAGCAGCGGG CCCCGGGCGC CGCCGCCGCG ATCCCTCCCC
 101 GCGCCCGCCG AGCACATCGC CGCCGCCGAG ATGGGCCCTC CGCGGCACCC
 151 CCAGGCCGGC GAGATAGAAG CGGGCGGTGC GGGCGGCGGG CGGCGGCTAC
 201 AGGTGGAAAT GAGTTCTCAA CAGTTTCCTC GGTTAGGAGC CCCTTCTACC
 251 GGGCTGAGCC AGGCCCCTTC TCAGATTGCA AACAGTGGTT CTGCTGGATT
 301 GATAAACCCA GCTGCTACAG TCAATGATGA ATCTGGTCGA GATTCTGAAG
 351 TCAGTGCCAG GGAGCACATG AGTTCCAGCA GCTCCCTCCA GTCCCGGGAG
 401 GAGAAGCAAG AGCCTGTTGT GGTAAGGCCC TATCCACAGG TGCAGATGTT
 451 GTCGACACAC CATGCTGTCG CATCAGCCAC ACCTGTTGCA GTGACAGCCC
 501 CGCCAGCACA CCTGACGCCA GCAGTGCCAC TTTCATTTTC GGAGGGACTT
 551 ATGAAGCCGC CCCCGAAGCC CACCATGCCT AGCCGTCCCA TTGCTCCTGC
 601 TCCACCTTCT ACCCTGTCAC TTCCCCCCAA GGTTCCAGGG CAGGTTACCG
 651 TTACCATGGA GAGTAGCATC CCTCAAGCTT CAGCCATTCC TGTGGCAACA
 701 ATCAGTGGAC AACAGGGCCA TCCCAGTAAC CTGCATCACA TCATGACTAC
 751 AAATGTGCAA ATGTCTATCA TCCGCAGCAA TGCTCCTGGG CCCCCTCTTC
 801 ACATTGGAGC TTCTCATTTA CCTCGAGGTG CAGCTGCTGC TGCTGTGATG
 851 TCCAGTTCTA AAGTAACCAC AGTCCTGAGG CCGACCTCAC AGCTGCCAAA
 901 TGCTGCTACT GCTCAGCCAG CAGTACAGCA CATCATTCAC CAACCAATCC
 951 AGTCTCGGCC ACCTGTGACC ACCTCCAATG CCATCCCTCC TGCTGTGGTA
1001 GCAACTGTCT CAGCCACCAG AGCTCAGTCT CCAGTCATCA CTACGACAGC
1051 GGCGCATGCT ACTGATTCAG CACTTAGTAG GCCAACCTTG TCTATCCAGC
1101 ATCCTCCATC TGCAGCAATC AGTATTCAGC GTCCTGCCCA GTCACGAGAT
1151 GTCACAACAA GAATCACACT ACCATCTCAC CCTGCATTAG GGACGCCAAA
1201 ACAGCAGCTT CATACAATGG CTCAGAAAAC AATCTTCAGT ACTGGCACGC
1251 CAGTGGCTGC AGCCACAGTA GCACCTATTT TGGCAACCAA CACCATTCCT
1301 TCAGCGACCA CAGCTGGATC TGTGTCACAC ACGCAAGCTC CCACAAGTAC
1351 CATTGTTACC ATGACAGTAC CCTCCCATTC CTCCCATGCT ACTGCTGTGA
1401 CCACCTCAAA CATCCCAGTC GCCAAGGTGG TGCCCCAGCA GATCACGCAC
1451 ACTTCTCCTC GGATCCAGCC AGACTACCCT GCCGAGAGGA GTAGCCTGAT
1501 TCCCATCTCC GGACATCGGG CCTCTCCCAA TCCTGTGGCC ATGGAAACCC
1551 GAAGTGACAA CAGACCGTCT GTTCCCGTTC AGTTCCAATA TTTTTTGCCA
1601 ACTTACCCCC CTTCTGCATA CCCACTGGCG GCACATACCT ACACCCCAAT
1651 CACCAGTTCC GTGTCCACTA TCCGACAGTA TCCAGTTTCA GCTCAGGCTC
1701 CAAACTCTGC CATCACAGCT CAGACTGGTG TTGGGGTAGC GTCTACCGTC
1751 CACCTAAACC CCATGCAGTT GATGACAGTG GATGCATCGC ATGCTCGACA
1801 TATTCAAGGG ATCCAGCCAG CACCCATCAG TACCCAGGGT ATCCAGCCGG
1851 CCCCCATTGG GACCCCAGGG ATACAGCCTG CACCACTTGG CACACAGGGA
1901 ATTCACTCAG CAACCCCAAT CAACACACAA GGGCTTCAGC CTGCACCTAT
1951 GGGTACTCAG CAGCCTCAGC CTGAAGGAAA GACTTCAGCA GTGGTGTTGG
2001 CAGATGGAGC CACAATTGTG GCCAACCCTA TTAGCAATCC ATTCAGTGCT
2051 GCTCCAGCAG CAACAACCGT GGTGCAGACC CACAGCCAGA GTGCTAGCAC
2101 CAACGCTCCC GCCCAGGGCT CATCGCCACG GCCAAGCATA CTCCGGAAGA
2151 AACCTGCCAC AGATGGTGCC AAACCCAAGT CTGAAATCCA CGTGTCTATG
2201 GCCACTCCGG TCACTGTGTC CATGGAGACT GTATCCAATC AAAATAATGA
2251 TCAGCCTACC ATTGCCGTCC CTCCAACTGC CCAGCAGCCC CCACCGACCA
2301 TTCCAACTAT GATTGCAGCA GCCAGTCCCC CGTCACAACC AGCCGTTGCC
2351 CTTTCAACCA TTCCTGGAGC GGTCCCCATC ACTCCACCCA TCACCACCAT
2401 TGCAGCTGCA CCACCTCCAT CAGTCACTGT GGGTGGCAGT CTTTCCTCCG
2451 TCTTGGGCCC TCCCGTTCCT GAAATTAAAG TGAAAGAAGA AGTAGAACCA
2501 ATGGATATCA TGAGGCCAGT TTCTGCAGTT CCTCCACTGG CTACCAACAC
2551 TGTGTCTCCA TCTCTTGCAT TGCTGGCAAA CAACTTGTCC ATGCCTACAA
2601 GTGACCTACC ACCTGGTGCC TCCCCAAGGA AAAAGCCTCG AAAGCAACAG
2651 CATGTGATCT CAACAGAAGA AGGTGACATG ATGGAGACAA ACAGCACTGA
2701 TGATGAGAAG TCCACTGCCA AGAGTCTTCT GGTGAAGGCT GAGAAGCGCA
```

```
2751 AGTCTCCTCC CAAGGAGTAT ATTGATGAGG AAGGTGTGAG ATATGTCCCA
2801 GTGCGTCCAA GACCCCCCAT TACTTTGCTT CGTCACTATC GGAACCCCTG
2851 GAAAGCTGCT TACCACCACT TTCAGAGGTA CAGTGACGTC CGGGTCAAAG
2901 AGGAGAAGAA AGCTATGCTG CAGGAAATAG CTAATCAGAA AGGAGTATCC
2951 TGTCGTGCTC AAGGCTGGAA AGTCCACCTC TGTGCTGCCC AGTTACTACA
3001 GCTGACGAAT CTAGAACATG ATGTCTATGA AAGACTTACT AACCTGCAGG
3051 AAGGGATTAT CCCAAAGAAA AAAGCAGCAA CAGATGATGA TCTCCACCGA
3101 ATAAACGAAC TGATACAGGG AAATATGCAG AGGTGTAAAC TTGTGATGGA
3151 TCAAATCAGT GAAGCCAGAG ACTCCATGCT TAAGGTTTTA GATCATAAAG
3201 ACCGTGTCCT GAAGCTGCTT AACAAGAACG GGACTGTCAA AAAAGTGTCC
3251 AAATTGAAGC GAAAGGAAAA AGTCTAGACC CAGAACAATC AGGAGATTGG
3301 AAGCAAATTT ATGAAGAATG ATGGTGGGGG TGGGGGGAGG GTTTTGGTTT
3351 TTTCCAAAGT GGAACATTGA AATAAAGGAA GTGTTCCTTA GTTCCCGTGT
3401 GAAAGCAGAG GAACCCATGA CATCCAAGGG CGTGAAAGGA TCAGAGCTGA
3451 CTGGACATAG TGAGCTGCCT TCTTGCGTTC GGGTGCACCC CTGTTAAACC
3501 TGATCTGTGT CATAAGTGAC TCCGGATGCA TCAGTGTCCA CCAGTTGGAA
3551 GCAATGACAA GGATGGCTGG CTGGTGTTTT TCAGCCTTCC GGTTTATAGA
3601 CTGTATTTAT CTAGTGGATT CCTGCAGGCC CCATACTGAG CCTGGACTGA
3651 AAGTATCCAC TCGGACCATC TGTTATCTCT CTACACTGAA AATAAAACCT
3701 CTTCCACCCA CCCCATTCGG TTCTTCTGCC TGACCTTCAA ATGCCCATGT
3751 TGGCCTTTTA CAGCAGTGCC ACGGCACCAA GCGAGCTGCC ACATCTCACA
3801 CTCTAAAGGG TTTGAACTAT TAGTTCTTGT CATTTTTTAA AAAAAACCAT
3851 TCCCAAGTGA AATTGTTATA TCGTCTGTCT TGCGTGTGTC AGAACTGGGT
3901 TTTTGTGGAG GTTCAGAGCA GGCAACACCA TAAGTTGCTC TCAGATCCTT
3951 GTTCTGAAGT ACATTCTTGG TTATCTGTAC TTCTGTAGCT GGTGTGATGC
4001 TGTTAATTGT ATGTACCACA CATCTCCAGA CGTTAATAAA GGACTCAAAG
4051 AGGTTTTTGT AAAAAAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 131 bp to 3274 bp; peptide length: 1048
Category: similarity to known protein

```
   1 MGPPRHPQAG EIEAGGAGGG RRLQVEMSSQ QFPRLGAPST GLSQAPSQIA
  51 NSGSAGLINP AATVNDESGR DSEVSAREHM SSSSSLQSRE EKQEPVVVRP
 101 YPQVQMLSTH HAVASATPVA VTAPPAHLTP AVPLSFSEGL MKPPPKPTMP
 151 SRPIAPAPPS TLSLPPKVPG QVTVTMESSI PQASAIPVAT ISGQQGHPSN
 201 LHHIMTTNVQ MSIIRSNAPG PPLHIGASHL PRGAAAAAVM SSSKVTTVLR
 251 PTSQLPNAAT AQPAVQHIIH QPIQSRPPVT TSNAIPPAVV ATVSATRAQS
 301 PVITTTAAHA TDSALSRPTL SIQHPPSAAI SIQRPAQSRD VTTRITLPSH
 351 PALGTPKQQL HTMAQKTIFS TGTPVAAATV APILATNTIP SATTAGSVSH
 401 TQAPTSTIVT MTVPSHSSHA TAVTTSNIPV AKVVPQQITH TSPRIQPDYP
 451 AERSSLIPIS GHRASPNPVA METRSDNRPS VPVQFQYFLP TYPPSAYPLA
 501 AHTYTPITSS VSTIRQYPVS AQAPNSAITA QTGVGVASTV HLNPMQLMTV
 551 DASHARHIQG IQPAPISTQG IQPAPIGTPG IQPAPLGTQG IHSATPINTQ
 601 GLQPAPMGTQ QPQPEGKTSA VVLADGATIV ANPISNPFSA APAATTVVQT
 651 HSQSASTNAP AQGSSPRPSI LRKKPATDGA KPKSEIHVSM ATPVTVSMET
 701 VSNQNNDQPT IAVPPTAQQP PPTIPTMIAA ASPPSQPAVA LSTIPGAVPI
 751 TPPITTIAAA PPPSVTVGGS LSSVLGPPVP EIKVKEEVEP MDIMRPVSAV
 801 PPLATNTVSP SLALLANNLS MPTSDLPPGA SPRKKPRKQQ HVISTEEGDM
 851 METNSTDDEK STAKSLLVKA EKRKSPPKEY IDEEGVRYVP VRPRPPITLL
 901 RHYRNPWKAA YHHFQRYSDV RVKEEKKAML QEIANQKGVS CRAQGWKVHL
 951 CAAQLLQLTN LEHDVYERLT NLQEGIIPKK KAATDDDLHR INELIQGNMQ
1001 RCKLVMDQIS EARDSMLKVL DHKDRVLKLL NKNGTVKKVS KLKRKEKV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2all, frame 2

SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2)., N = 1,
Score = 334, P = 2.4e-25

PIR:A43932 mucin 2 precursor, intestinal - human (fragments), N = 1, Score = 321, P = 3.2e-24

TREMBL:D88440_1 product: "high molecular mass nuclear antigen"; Gallus gallus mRNA for high molecular mass nuclear antigen, partial cds., N = 1, Score = 312, P = 8.3e-24

PIR:S48478 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae), N = 1, Score = 300, P = 2.1e-22

>SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
Length = 5,179

HSPs:

Score = 334 (50.1 bits), Expect = 2.4e-25, P = 2.4e-25
Identities = 184/770 (23%), Positives = 263/770 (34%)

```
Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V   P P       T    + + T V   T  P     TP         +     + P P PT      P
Sbjct:  3471 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 3530

Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
               P   +T   P   P G  T T          + P   T +G Q  P+        TT V  +
Sbjct:  3531 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 3589

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
               +   P   P+   +   P                 +++ +TT     T        T  P    I
Sbjct:  3590 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 3649

Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
                   + P  T      P      T + T    +P  T T        T      +  T++      P
Sbjct:  3650 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTVT---PTPT 3706

Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
                   Q P   +     TT        P+        GT            +       T +     T TP       T           PI
Sbjct:  3707 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 3766

Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T  G+ + T  P +T  T+T  P+ +    T  TT           V P    T T
Sbjct:  3767 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 3825

Query:   444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
             +      P    ++ +          +P P    +T +      + P+       + PT  P+
Sbjct:  3826 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 3874

Query:   503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
             T TPIT++ + T      P    Q P +   IT  T V        T        Q  T
Sbjct:  3875 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 3932

Query:   561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
               P P   TQ         PI T       P P GTQ    + TPI T         P P GTQ P
Sbjct:  3933 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTT 3991

Query:   614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
             P     T+ V        T        P + P +          T T       T +Q+ +T           ++   P+
Sbjct:  3992 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT 4051

Query:   672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMI 728
                   T      P +           TP        +T +              T         P          PT   Q P T P
Sbjct:  4052 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 4111

Query:   729 AAASPPSQPAVALSTIPGAVPITPPITTIAAAPPPS-----VTVGGSLSSVLGP-PVPEI 782
                   P+              T P      PIT       TT+      P P+            T           +  ++ +   P P P
Sbjct:  4112 TTTVTPTPTPTGTQT-PTTTPITTT-TTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG 4169

Query:   783 KVKEEVEPMDIMRPVSAVP-PLATNTVSPSLALLANNLSMPTSDLPPGASPRKKPRKQQH 841
                   P+        V+    P P    T T    P+          A           +  TS+    PP  +S      +     R
Sbjct:  4170 TQTPTTTPITTTTTVTPTPTPTGTQTGPPTHTSTAPIAELTTSNPPPESSTPQTSRSTSS 4229

Query:   842 VISTEEGDMMET 853
               + TE     ++ T
Sbjct:  4230 PL-TESTTLLST 4240
```

Score = 328 (49.2 bits), Expect = 1.0e-24, P = 1.0e-24
Identities = 180/745 (24%), Positives = 254/745 (34%)

```
Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V   P P       T    + + T V   T  P     TP         +     + P P PT      P
Sbjct:  3540 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 3599
```

```
Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
             P  +T  P   P G  T T          +  P  T +G Q  P+       TT V  +
Sbjct:  3600 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 3658

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
              +     P   P+      + P            +++ +TT      T          T  P    I
Sbjct:  3659 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 3718

Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
               + P  T     P       T + T    +P  T T      T     +  T++     P
Sbjct:  3719 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 3775

Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
              Q P  +     TT      P+      GT           +     T +  T TP      T      PI
Sbjct:  3776 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 3835

Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T   G+ + T   P +T   T+T   P+ +      T   TT            V P     T T
Sbjct:  3836 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 3894

Query:   444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
             +        P    ++ +          +P P     +T +        + P+        + PT   P+
Sbjct:  3895 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 3943

Query:   503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
             T TPIT++ + T      P     Q P +  IT  T V       T           Q  T
Sbjct:  3944 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 4001

Query:   561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
               P P  TQ      PI T          P P GTQ      + TPI T             P P GTQ P
Sbjct:  4002 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTT 4060

Query:   614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
              P   T+ V          T      P + P +          T T      T +Q+ +T             ++  P+
Sbjct:  4061 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT 4120

Query:   672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMI 728
                   T     P +              TP          +T +           T     P     PT  Q P T P
Sbjct:  4121 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 4180

Query:   729 AAASPPSQPAVALSTIPGAVPITPPITTIAAA-PPPSVTVGGSLSSVLGPPVPEIKVKEE 787
                   P+             T P           T PI   +   + PPP   +     + S       P        +
Sbjct:  4181 TTTVTPTPTPTGTQTGPPTHTSTAPIAELTTSNPPPESSTPQTSRSTSSPLTESTTLLST 4240

Query:   788 VEPMDIMRPVSAVPPLATNTVSPSLALLANNLSMP--TSDLPPGASPR 833
             + P     M     S   PP +T T +P+        + LS P   T+   PPG     R
Sbjct:  4241 LPPAIEM--TSTAPP-STPT-APTTTSGGHTLSPPPSTTTSPPGTPTR 4284
```

Score = 325 (48.8 bits), Expect = 2.2e-24, P = 2.2e-24
Identities = 186/782 (23%), Positives = 261/782 (33%)

```
Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V    P P          T       + + T V   T  P        TP                  +     + P P PT         P
Sbjct:  3494 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 3553

Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
             P  +T  P   P G  T T          +  P  T +G Q  P+       TT V  +
Sbjct:  3554 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 3612

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
              +     P   P+      + P            +++ +TT      T          T  P    I
Sbjct:  3613 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 3672

Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
               + P  T     P       T + T    +P  T T      T     +  T++     P
Sbjct:  3673 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 3729

Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
              Q P  +     TT      P+      GT           +     T +  T TP      T      PI
Sbjct:  3730 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 3789

Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T   G+ + T   P +T   T+T   P+ +      T   TT            V P     T T
Sbjct:  3790 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 3848

Query:   444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
             +        P    ++ +          +P P     +T +        + P+        + PT   P+
Sbjct:  3849 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 3897

Query:   503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
```

```
                 T TPIT++ + T     P   Q P +  IT   T V     T       Q  T
Sbjct:  3898 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 3955

Query:   561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
               P P  TQ       PI T       P P GTQ    + TPI T       P P GTQ P
Sbjct:  3956 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTT 4014

Query:   614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
              P   T+ V       T      P + P +      T T    T +Q+ +T       ++  P+
Sbjct:  4015 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT 4074

Query:   672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMI 728
                  T    P +       TP       +T +        T    P    PT  Q P T P
Sbjct:  4075 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 4134

Query:   729 AAASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
                P+         T P    PIT    TT      P P+ T   G+ +      P      I       V
Sbjct:  4135 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQT----PTTTPITTTTTV 4184

Query:   789 EPMDIMRPVSAVPPLATNTVSPSLALLANNLSMPTSDLPPGASPRKKPRKQQHVISTEEG 848
               P           PP  T+T +P    L  +N   P S  P  +       P  +    + +
Sbjct:  4185 TPTPTPTGTQTGPPTHTST-APIAELTTSN-PPPESSTPQTSRSTSSPLTESTTLLSTLP 4242

Query:   849 DMMETNSTDDEKSTAKSLLVKAEKRKSPP 877
               +E  ST      +               SPP
Sbjct:  4243 PAIEMTSTAPPSTPTAPTTTSGGHTLSPP 4271

 Score = 324 (48.6 bits), Expect = 2.8e-24, P = 2.8e-24
 Identities = 170/717 (23%), Positives = 248/717 (34%)

Query:    95 PVVVRPYPQVQMLSTHHAVASATP--VAVTAPPAHLTPAVPLSFSEGLMKPPPKPTMPSR 152
             P     P P      +T        + +P       T PP     TP+  P++ +     + P      P+ P
Sbjct:  1401 PPTTTPSPPPTTTTTLPPTTTPSPPTTTTTTPPPTTTPSPPITTTTTPL-PTTTPSPPIS 1459

Query:   153 PIAPAPPSTLSLPPKVPGQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
                   PP+T     PP            T  S        + P  T +       P          I            +
Sbjct:  1460 TTTTPPPTTTPSPPTTTPSPPTTTPSPPTTTTTTPPPTTTPS---PPMTTPITPPASTTT 1516

Query:   213 IIRSNAPGPPLHIGASHLPRGAAAAAVMSSSKVTTVLRPTSQ--LPNAATAQPAVQHIIH 270
             +   +  P PP       +  P              S      T +  PTS    LP    T    P
Sbjct:  1517 LPPTTTPSPPTTTTTTPPP-----TTTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTT 1571

Query:   271 QPIQSRP-PVTTSNAIPPAVVATVSA-TRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
              P  + P P TT+      PP  +   T       T       SP    TTT          + S PT +      PP++
Sbjct:  1572 PPPTTTPSPPTTTTPSPPTITTTTPPPTTTPSPPTTTTTPPPTTTPSPPTTTPITPPTS 1631

Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPILATNT 388
               ++                T    T  P  P    TP         T         I  +T  TP      T  +   + T
Sbjct:  1632 TTTLPPTTTPSPPPTTTTTP--PPTTTPSPPTTTTPSPPITTTTTPPPTTTPSSPITTTP 1689

Query:   389 IPSATTAGSVSHTQAPTSTIVTMTVPSHSSHATAV-TTSNIPVAKVVPQQITHTSPRIQP 447
              P  TT   + S T   P+S I T T PS ++   +     TT      P      P      T T     + P
Sbjct:  1690 SPPTTTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPTTMTTPSPTTTPSPPTTTMTTLPP 1749

Query:   448 DYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPV-QFQYFLPTYPPSAY-P-----LA 500
                +     +  P+             P             T  +      P VP+    +      +L  +      P+ + P          L
Sbjct:  1750 TTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTELI 1809

Query:   501 AHTYTPITSSVSTIR--QYP-VSAQAPNSAITAQTGVG-VASTVHLNPMQLMTVDASHAR 556
                  P  ++   + R      YP  V                +          VG  +             P   ++  + A
Sbjct:  1810 GDVCGPGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPM-AFCLN 1868

Query:   557 HIQGIQPAPISTQGIQPAPIGTPGIQ-PAPLGTQGIHSATPINTQGLQPAPMGTQQPQ-- 613
             +       +Q          TQ       P   + T      + P P      T      I    + T  +                P P GTQ P
Sbjct:  1869 YEINVQCCECVTQ---PTTMTTTTENPTPPTTTPITTTTTVTPT---PTPTGTQTPTTT 1922

Query:   614 PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSILR 672
             P     T+ V        T        P + P +          T T      T +Q+ +T            ++  P+
Sbjct:  1923 PITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP 1982

Query:   673 KKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMIA 729
                 T    P +         TP         +T +          T     P     PT   Q P T P
Sbjct:  1983 TGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTT 2042

Query:   730 AASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEVE 789
               P+          T P    PIT    TT       P P+ T   G+ +         P          V
Sbjct:  2043 TTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTPT 2096

Query:   790 PMDIMRPVSAVPPLATNTVSPS 811
             P     P +   P       T TV+P+
Sbjct:  2097 PTGTQTPTTT-PITTTTTVTPT 2117
```

```
 Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V   P P          T    + + T V  T  P     TP         +    + P P PT       P
Sbjct:  2068 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 2127

Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
               P   +T   P   P G  T T           +  P  T +G Q  P+        TT V  +
Sbjct:  2128 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 2186

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
               +    P   P+      +  P          +++ +TT       T        T  P   I
Sbjct:  2187 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2246

Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
                 +  P  T      P       T + T    +P  T T      T       +  T++    P
Sbjct:  2247 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 2303

Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
                 Q P   +      TT       P+      GT            +   T +  T TP      T     PI
Sbjct:  2304 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 2363

Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T   G+ + T  P +T  T+T  P+ +       T  TT         V P   T T
Sbjct:  2364 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 2422

Query:   444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
             +      P     ++ +          +P P     +T +        + P+         + PT   P+
Sbjct:  2423 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 2471

Query:   503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
             T TPIT++ + T      P    Q P +   IT   T V        T          Q  T
Sbjct:  2472 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 2529

Query:   561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
               P P  TQ         PI T        P P GTQ     + TPI T          P P GTQ P
Sbjct:  2530 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTT 2588

Query:   614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
              P    T+ V       T        P + P +          T  T      T +Q+ +T            ++   P+
Sbjct:  2589 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT 2648

Query:   672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMI 728
                  T     P +            TP          +T +           T     P      PT   Q P T P
Sbjct:  2649 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 2708

Query:   729 AAASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
                  P+           T P      PIT       TT         P P+ T   G+ +          P              V
Sbjct:  2709 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTP 2762

Query:   789 EPMDIMRPVSAVPPLATNTVSPS 811
              P       P  +  P       T TV+P+
Sbjct:  2763 TPTGTQTPTTT-PITTTTTVTPT 2784
```



```
 Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V   P P          T    + + T V  T  P     TP         +    + P P PT       P
```



```
Sbjct:  2206 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 2265
```



```
Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
               P   +T   P   P G  T T           +  P  T +G Q  P+        TT V  +
```



```
Sbjct:  2266 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 2324
```



```
Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
               +    P   P+      +  P          +++ +TT       T        T  P   I
```



```
Sbjct:  2325 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2384
```



```
Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
                 +  P  T      P       T + T    +P  T T      T       +  T++    P
```



```
Sbjct:  2385 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 2441
```



```
Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
                 Q P   +      TT       P+      GT            +   T +  T TP      T     PI
```



```
Sbjct:  2442 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 2501
```



```
Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T   G+ + T  P +T  T+T  P+ +       T  TT         V P   T T
```

```
Sbjct: 2502 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 2560

Query:  444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
            +      P   ++ +        +P P    +T +      + P+       + PT   P+
Sbjct: 2561 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 2609

Query:  503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
            T TPIT++ + T     P   Q P +  IT  T V     T         Q  T
Sbjct: 2610 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 2667

Query:  561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
              P P  TQ       PI T         P P GTQ    + TPI T          P P GTQ P
Sbjct: 2668 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTT 2726

Query:  614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
             P    T+ V         T      P + P +         T T    T +Q+ +T         ++   P+
Sbjct: 2727 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT 2786

Query:  672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMI 728
                    T     P +           TP        +T +            T     P      PT   Q  P  T  P
Sbjct: 2787 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 2846

Query:  729 AAASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
                   P+            T P     PIT      TT        P P+ T    G+ +        P            V
Sbjct: 2847 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTP 2900

Query:  789 EPMDIMRPVSAVPPLATNTVSPS 811
             P        P +   P       T TV+P+
Sbjct: 2901 TPTGTQTPTTT-PITTTTTVTPT 2922
```

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

```
Query:   96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
            V     P P          T     + + T V   T   P        TP               +       + P P PT       P
Sbjct: 2321 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 2380

Query:  155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
              P   +T    P      P G  T T                      + P  T +G Q  P+              TT V  +
Sbjct: 2381 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 2439

Query:  213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
               +    P     P+     + P                  +++ +TT       T             T   P   I
Sbjct: 2440 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2499

Query:  269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
                  +  P  T        P         T + T      +P  T T       T             +  T++     P
Sbjct: 2500 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 2556

Query:  329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
                Q P   +      TT       P+        GT             +    T +  T TP       T      PI
Sbjct: 2557 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 2616

Query:  386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T   G+ + T   P +T  T+T  P+ +      T   TT            V P    T T
Sbjct: 2617 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 2675

Query:  444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
            +      P   ++ +        +P P    +T +      + P+       + PT   P+
Sbjct: 2676 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 2724

Query:  503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
            T TPIT++ + T     P   Q P +  IT  T V     T         Q  T
Sbjct: 2725 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 2782

Query:  561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
              P P  TQ       PI T         P P GTQ    + TPI T          P P GTQ P
Sbjct: 2783 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTT 2841

Query:  614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
             P    T+ V         T      P + P +         T T    T +Q+ +T         ++   P+
Sbjct: 2842 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT 2901

Query:  672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMI 728
                    T     P +           TP        +T +            T     P      PT   Q  P  T  P
Sbjct: 2902 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 2961

Query:  729 AAASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
                   P+            T P     PIT      TT        P P+ T    G+ +        P            V
Sbjct: 2962 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTP 3015
```

```
Query:   789 EPMDIMRPVSAVPPLATNTVSPS 811
             P     P +  P    T TV+P+
Sbjct:  3016 TPTGTQTPTTT-PITTTTTVTPT 3037

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V   P P        T    + + T V   T  P      TP         +     + P P PT     P
Sbjct:  2390 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 2449

Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
               P   +T   P   P G  T T             +  P  T +G Q  P+         TT V  +
Sbjct:  2450 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 2508

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
               +    P   P+     + P             +++ +TT     T          T  P     I
Sbjct:  2509 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2568

Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
                   + P T      P      T + T     +P  T T      T       +  T++      P
Sbjct:  2569 TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 2625

Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
                  Q P   +    TT     P+     GT             +     T +  T TP      T        PI
Sbjct:  2626 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 2685

Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T   G+ + T   P +T  T+T  P+ +    T   TT           V P    T T
Sbjct:  2686 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 2744

Query:   444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
             +      P     ++ +         +P P      +T +       + P+         + PT    P+
Sbjct:  2745 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 2793

Query:   503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
             T TPIT++ + T      P     Q P +   IT   T V       T             Q  T
Sbjct:  2794 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 2851

Query:   561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
               P P  TQ       PI T         P P GTQ    + TPI T             P P GTQ P
Sbjct:  2852 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTT 2910

Query:   614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
              P    T+ V         T      P + P +         T T       T +Q+ +T           ++     P+
Sbjct:  2911 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT 2970

Query:   672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMI 728
                 T      P +           TP         +T +             T      P      PT   Q P T P
Sbjct:  2971 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 3030

Query:   729 AAASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
                  P+        T P     PIT       TT        P P+ T   G+ +          P              V
Sbjct:  3031 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTP 3084

Query:   789 EPMDIMRPVSAVPPLATNTVSPS 811
             P     P +  P    T TV+P+
Sbjct:  3085 TPTGTQTPTTT-PITTTTTVTPT 3106

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V   P P        T    + + T V   T  P      TP         +     + P P PT     P
Sbjct:  2459 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 2518

Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
               P   +T   P   P G  T T             +  P  T +G Q  P+         TT V  +
Sbjct:  2519 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 2577

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
               +    P   P+     + P             +++ +TT     T          T  P     I
Sbjct:  2578 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2637

Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
                   + P T      P      T + T     +P  T T      T       +  T++      P
Sbjct:  2638 TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 2694

Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
                  Q P   +    TT     P+     GT             +     T +  T TP      T        PI
Sbjct:  2695 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 2754
```

```
Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T  G+ + T  P +T  T+T  P+ +    T  TT         V P   T T
Sbjct:  2755 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 2813

Query:   444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
             +      P   ++ +          +P P   +T +    + P+      + PT  P+
Sbjct:  2814 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 2862

Query:   503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
             T TPIT++ + T     P    Q P +  IT  T V     T         Q  T
Sbjct:  2863 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 2920

Query:   561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
               P P  TQ       PI T        P P GTQ    + TPI T      P P GTQ P
Sbjct:  2921 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTT 2979

Query:   614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
              P   T+ V        T     P + P +       T T    T +Q+ +T       ++  P+
Sbjct:  2980 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT 3039

Query:   672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMI 728
                 T      P +          TP         +T +          T   P    PT  Q P T P
Sbjct:  3040 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 3099

Query:   729 AAASPPSPQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
                P+            T P     PIT     TT        P P+ T  G+ +        P           V
Sbjct:  3100 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTP 3153

Query:   789 EPMDIMRPVSAVPPLATNTVSPS 811
             P      P +  P     T TV+P+
Sbjct:  3154 TPTGTQTPTTT-PITTTTTVTPT 3175
```

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
Identities = 174/717 (24%), Positives = 243/717 (33%)

```
Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V    P P              T     + + T  V  T  P     TP          +     + P P PT        P
Sbjct:  2528 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 2587

Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
              P  +T    P   P  G  T  T             +  P  T +G Q P+       TT V  +
Sbjct:  2588 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 2646

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
             +      P    P+      +  P                    +++ +TT        T            T   P     I
Sbjct:  2647 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2706

Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
                 + P   T        P            T + T    +P  T T        T         +   T++ +      P
Sbjct:  2707 TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 2763

Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
                Q P   +    TT        P+      GT           +      T +  T TP       T       PI
Sbjct:  2764 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 2823

Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T  G+ + T  P +T  T+T  P+ +    T  TT         V P   T T
Sbjct:  2824 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 2882

Query:   444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
             +      P   ++ +          +P P   +T +    + P+      + PT  P+
Sbjct:  2883 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 2931

Query:   503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
             T TPIT++ + T     P    Q P +  IT  T V     T         Q  T
Sbjct:  2932 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 2989

Query:   561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
               P P  TQ       PI T        P P GTQ    + TPI T      P P GTQ P
Sbjct:  2990 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTT 3048

Query:   614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
              P   T+ V        T     P + P +       T T    T +Q+ +T       ++  P+
Sbjct:  3049 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT 3108

Query:   672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMI 728
                 T      P +          TP         +T +          T   P    PT  Q P T P
Sbjct:  3109 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 3168

Query:   729 AAASPPSPQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
```

```
                   P+          T P    PIT    TT     P P+ T  G+ +     P        V
Sbjct:  3169 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTP 3222

Query:   789 EPMDIMRPVSAVPPLATNTVSPS 811
               P     P + P    T TV+P+
Sbjct:  3223 TPTGTQTPTTT-PITTTTTVTPT 3244

 Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V   P P        T     + + T V  T  P     TP         +     + P P PT      P
Sbjct:  3080 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 3139

Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
               P   +T    P    P G  T T             +  P  T +G Q  P+       TT V  +
Sbjct:  3140 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 3198

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
                +     P   P+     + P                +++ +TT      T       T P     I
Sbjct:  3199 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 3258

Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
                   + P T      P       T + T    +P  T T      T       +   T++ +     P
Sbjct:  3259 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 3315

Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
                 Q P  +   TT     P+    GT        +   T +  T TP     T     PI
Sbjct:  3316 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 3375

Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T  G+ + T  P +T  T+T  P+ +   T   TT          V P   T T
Sbjct:  3376 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 3434

Query:   444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
             +      P   ++ +         +P P    +T +    + P+      + PT   P+
Sbjct:  3435 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 3483

Query:   503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
             T TPIT++ + T        P    Q P +  IT  T V     T          Q  T
Sbjct:  3484 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 3541

Query:   561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613
               P P   TQ         PI T       P P GTQ     + TPI T        P P GTQ P
Sbjct:  3542 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTT 3600

Query:   614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671
              P   T+ V           T    P + P +      T T     T +Q+ +T        ++  P+
Sbjct:  3601 TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT 3660

Query:   672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMI 728
                   T       P +        TP       +T +          T     P    PT  Q P T P
Sbjct:  3661 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 3720

Query:   729 AAASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
                   P+          T P    PIT    TT      P P+ T  G+ +      P        V
Sbjct:  3721 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTP 3774

Query:   789 EPMDIMRPVSAVPPLATNTVSPS 811
               P     P + P    T TV+P+
Sbjct:  3775 TPTGTQTPTTT-PITTTTTVTPT 3796

 Score = 313 (47.0 bits), Expect = 4.2e-23, P = 4.2e-23
 Identities = 169/695 (24%), Positives = 245/695 (35%)

Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V   P P        T     + + T V  T  P     TP         +     + P P PT      P
Sbjct:  3655 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 3714

Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
               P   +T    P    P G  T T             +  P  T +G Q  P+       TT V  +
Sbjct:  3715 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 3773

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
                +     P   P+     + P                +++ +TT      T       T P     I
Sbjct:  3774 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 3833

Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
                   + P T      P       T + T    +P  T T      T       +   T++ +     P
Sbjct:  3834 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 3890
```

```
Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
                 Q P  +   TT    P+   GT         +  T +  T TP    T    PI
Sbjct:  3891 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 3950

Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T  G+ + T  P +T  T+T  P+ +   T  TT        V P   T T
Sbjct:  3951 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 4009

Query:   444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
             +     P   ++ +        +P P   +T +    + P+      + PT  P+
Sbjct:  4010 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 4058

Query:   503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
             T TPIT++ + T     P   Q P +  IT   T V     T        Q  T
Sbjct:  4059 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 4116

Query:   561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQP 614
               P P  TQ     PI T       P P GTQ   + TPI T       P P GTQ P
Sbjct:  4117 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPT- 4174

Query:   615 EGKTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSSPRPSILRKK 674
                T+ +         T+     P        P      T       ++  ++N P + S+P+ S
Sbjct:  4175 ---TTPITTT--TTVTPTPTPTGTQTGPPTHTSTAPIAELTTSNPPPESSTPQTSRSTSS 4229

Query:   675 PATDGAKPKSEIH--VSMATPVTVSMETVSNQNNDQPTIAVPP-TAQQPP--PTIPTMIA 729
             P T+       S +    + M +     S  T      +    T++ PP T    PP   PT   T
Sbjct:  4230 PLTESTTLLSTLPPAIEMTSTAPPSTPTAPTTTSGGHTLSPPPSTTTSPPGTPTRGTTTG 4289

Query:   730 AASPPSQPAVALSTI----PGAVPITPP--ITTIAAAP-PPSVTVGGSLSSVLGPPVPEI 782
             ++S P+    V  +T       P    P++ P  I T    P P SV +   L+       P  E+
Sbjct:  4290 SSSAPTPSTVQTTTTSAWTPTPTPLSTPSIIRTTGLRPYPSSVLICCVLNDTYYAPGEEV 4349
```

Score = 279 (41.9 bits), Expect = 1.8e-19, P = 1.8e-19
Identities = 138/540 (25%), Positives = 194/540 (35%)

```
Query:   278 PVTTSNAIPPAVVATVSATRAQSPVITTTAAH------ATDSALSRP--TLSIQHPPSAA 329
             P+TT+  + P     T + T    +P+ TTT                T +  + P  T +    P
Sbjct:  1946 PITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP 2005

Query:   330 ISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILAT 386
                Q P  +   TT    P+   GT         +  T +  T TP    T    PI  T
Sbjct:  2006 TGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTT 2065

Query:   387 NTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSPR 444
             T+ P+ T  G+ + T  P +T  T+T  P+ +   T  TT        V P   T T  +
Sbjct:  2066 TTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGTQ 2124

Query:   445 IQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAHT 503
                 P   ++ +        +P P   +T +    + P+      + PT  P+         T
Sbjct:  2125 TPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTPT 2173

Query:   504 YTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQGI 561
             TPIT++ + T     P   Q P +  IT   T V     T        Q  T
Sbjct:  2174 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVTP 2231

Query:   562 QPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ-- 613
             P P  TQ     PI T       P P GTQ   + TPI T       P P GTQ P
Sbjct:  2232 TPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTTT 2290

Query:   614 PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSILR 672
             P   T+ V     T    P + P +     T T     T +Q+ +T       ++   P+
Sbjct:  2291 PITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP 2350

Query:   673 KKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMIA 729
               T       P +       TP       +T +       T     P       PT    Q  P  T  P
Sbjct:  2351 TGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTT 2410

Query:   730 AASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEVE 789
                   P+       T P    PIT    TT     P P+ T   G+ +       P        V
Sbjct:  2411 TTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTPT 2464

Query:   790 PMDIMRPVSAVPPLATNTVSPS 811
             P     P +  P    T TV+P+
Sbjct:  2465 PTGTQTPTTT-PITTTTTVTPT 2485
```

Score = 265 (39.8 bits), Expect = 5.8e-18, P = 5.8e-18
Identities = 179/746 (23%), Positives = 257/746 (34%)

```
Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V    P P        T     + + T V T P    TP          +  + P P PT      P
Sbjct:  3678 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 3737
```

```
Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
             P   +T   P   P G  T T          +  P  T +G Q  P+      TT V  +
Sbjct:  3738 TTPITTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 3796

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
               +    P   P+     + P           +++ +TT     T         T  P   I
Sbjct:  3797 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 3856

Query:   269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328
                +  P  T     P       T + T   +P  T T     T     +  T++    P
Sbjct:  3857 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT---PTPT 3913

Query:   329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385
                 Q P   +    TT      P+      GT          +   T +  T TP     T     PI
Sbjct:  3914 PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT 3973

Query:   386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTSP 443
             T T+ P+ T  G+ + T  P +T  T+T  P+ +    T  TT          V P   T T
Sbjct:  3974 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP-TPTGT 4032

Query:   444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502
              +       P    ++ +          +P P    +T +      + P+      + PT  P+
Sbjct:  4033 QTPTTTPITTTTTVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTP 4081

Query:   503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560
             T TPIT++ + T       P    Q P +  IT  T V    T        Q  T
Sbjct:  4082 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVT 4139

Query:   561 IQPAPISTQGIQPAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQP 614
              P P  TQ       PI T        P P GTQ    + TPI T       P P GTQ   P
Sbjct:  4140 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTGPP 4198

Query:   615 EGKTSAVVLADGATIVANPISNPFSAAPA---ATTVVQTHSQSA-STNAPA--QGSSPRP 668
               TS    +A+   T   +NP    P S+ P        +T+      T S +   ST   PA     S+  P
Sbjct:  4199 T-HTSTAPIAELTT--SNP--PPESSTPQTSRSTSSPLTESTTLLSTLPPAIEMTSTAPP 4253

Query:   669 SILRKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVPPTAQQPPPTIPTMI 728
             S         T G    S       +  +P           +  ++  PT  +    T       T PT
Sbjct:  4254 STPTAPTTTSGGHTLSPPPSTTTSPPGTPTRGTTTGSSSAPTPSTVQTTTTSAWT-PTPT 4312

Query:   729 AAASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEV 788
              ++P           L     P +V  I     +        AP     V     G+            +    E
Sbjct:  4313 PLSTPSIIRTTGLRPYPSSVLICCVLNDTYYAPGEEV-YNGTYGDTCYFVNCSLSCTLEF 4371

Query:   789 EPMDIMRPVSAVPPLATNTVSPSLALLANNLSMPTSDLPPGASPRKKPRKQQH 841
                         S   P   +  +T +PS      ++   S PT     P     P     P +Q++
Sbjct:  4372 YNWSCPSTPSPTPTPSKSTPTPSKP--SSTPSKPTPGTKPPECPDFDPPRQEN 4422
```

Score = 254 (38.1 bits), Expect = 8.7e-17, P = 8.7e-17
Identities = 167/697 (23%), Positives = 245/697 (35%)

```
Query:   115 SATPVAVTAPPAHLTPAVPLSFSEGLMKPPPK--PTMPSR-PIAPAPPSTLSLPPKV-PG 170
             S     +   T  PP    TP+  P  +  +     PPP    P+  P+   PI  P  P  ST  +LPP     P
Sbjct:  1587 SPPTITTTTPPPTTTPSPPTTTTT---TPPPTTTPSPPTTTPITP-PTSTTTLPPTTTPS 1642

Query:   171 QVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMSIIRSNAPGPPLHIGASHL 230
                   T    +   P    +   P   T  +         +       TT          I    +   P PP          +
Sbjct:  1643 PPPTTTTTPPPTTTPSPPTTTTPSPPITTTTTPPPTTTPSSPI--TTTPSPPTTTMTTPS 1700

Query:   231 PRGAAAAAVMSSSKVTTVLRPTSQLPNAATAQPAVQHIIHQPIQS-RPPVTTSNAIPPAV 289
             P       SS  +TT     P+S       +  P            P  +   PP  TT       +PP
Sbjct:  1701 P------TTTPSSPITTTTTPSS---TTTPSPPPTTMTTPSPTTTPSPPTTMTTLPPTT 1751

Query:   290 VATVSATRAQSPVITT-TAAHATDSALSRPTLSIQH----PPSAAISIQRPAQSRDVTTR 344
              ++    T       P IT   T  +   +  +  +   P  +  +  +             S    +  +P       ++
Sbjct:  1752 TSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTELIGD 1811

Query:   345 ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPILATN-------TIPSATTAGS 397
              +   P     A    +  +  ++                I       G       V               ++   N                 IP A
Sbjct:  1812 VCGPGWAANISCRATMYP--DVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPMAFCLNY 1869

Query:   398 VSHTQAPTSTI--VTMTVPSHSSHATAVTTSNIPVAKVVPQQITHTSPRIQPDYPAERSS 455
              + Q            TMT   +  +  +  T   TT+  I           V        T  T     +          P    ++
Sbjct:  1870 EINVQCCECVTQPTTMTTTT-TENPTPPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTT 1928

Query:   456 LIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAHTYTPITSSVS-T 513
              +          +P P      +T +         + P+       + PT       P+                  T TPIT++ + T
Sbjct:  1929 TVT-----PTPTPTGTQTPT----TTPITTTTTVTPTPTPTG--TQTPTTTPITTTTTVT 1977

Query:   514 IRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQGIQPAPISTQGIQ 572
```

```
                P   Q P +  IT  T V     T       Q  T              P P  TQ
Sbjct:  1978 PTPTPTGTQTPTTTPITTTTTVTPTPTP--TGTQTPTTTPITTTTTVTPTPTPTGTQTPT 2035

Query:   573 PAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ--PEGKTSAVVLA 624
               PI T        P P GTQ    + TPI T        P P GTQ P    P    T+ V
Sbjct:  2036 TTPITTTTTVTPTPTPTGTQ-TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPT 2094

Query:   625 DGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSILRKKPATDGAKPK 683
               T     P + P +      T T    T +Q+ +T       ++   P+         T      P
Sbjct:  2095 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2154

Query:   684 SEIHVSMATPVTVSMETVSNQNNDQPTIAVP---PTAQQPPPTIPTMIAAASPPSQPAVA 740
             +          TP       +T +          T    P    PT  Q P T P           P+
Sbjct:  2155 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG 2214

Query:   741 LSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEVEPMDIMRPVSAV 800
               T P    PIT     TT      P P+ T   G+ +     P         V        P      P +
Sbjct:  2215 TQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTVTPTPTPTGTQTPTTT- 2267

Query:   801 PPLATNTVSPS 811
             P    T TV+P+
Sbjct:  2268 PITTTTTVTPT 2278

 Score = 243 (36.5 bits), Expect = 1.3e-15, P = 1.3e-15
 Identities = 110/406 (27%), Positives = 154/406 (37%)

Query:   121 VTAP-PAHLTPAVPLSFSEGLMKPPPKPTMPSRPIAPAPPSTLSLPPKVPGQVTVTMESS 179
             +T P P    TP+ P + +   L   P     P+ P+         PP+T     PP          T  + ++
Sbjct:  1396 ITTPSPPTTTPSPPPTTTTTL-PPTTTPSPPTTTTTTPPPTTTPSPPITT--TTTPLPTT 1452

Query:   180 IPQASAIPVATISGQQGHPSNLHHIMTTNVQMSIIRSNAPGPPLHIGASHLPRGAAAAAV 239
              P       P++T +        P+         TT                + P PP      + P
Sbjct:  1453 TPSP---PISTTTTPP--PTTTPSPPTTTPSPP---TTTPSPPTTTTTPPP------TT 1498

Query:   240 MSSSKVTTVLRP---TSQLPNAATAQPAVQHIIHQPIQSRP-PVTTSNAIPPAVVATVSA 295
               S  +TT + P     T+ LP    T  P              P  + P P TT+     PP      T+
Sbjct:  1499 TPSPPMTTPITPPASTTTLPPTTTPSPPTTTTTTPPPTTTPSPPTTTPITPPTSTTTLPP 1558

Query:   296 TRAQSPVITTTAAHATDSALSRPTLSIQHPPSAAISIQRPAQSRDV-TTRITLPSHPALG 354
              T    SP   TTT          +  S PT +    PP+     +     P   +          TT  T P  P
Sbjct:  1559 TTTPSPPPTTTTTPPPTTTPSPPTTTTPSPPTITTTTPPPTTTPSPPTTTTTTP--PPTT 1616

Query:   355 TPKQQLHTMAQKTIFSTGTPVAAATVAPILATNTIPSATTAGSVSHTQAPTSTIVTMTVP 414
             TP        T          +T   P      T +P      T T P    TT     S       T   P+  I T T P
Sbjct:  1617 TPSPPTTTPITPPTSTTTLP-PTTTPSPPPTTTTTPPPTTTPSPPTTTPSPPITTTTTP 1675

Query:   415 SHSSHATA-VTTSNIPVAKVVPQQITHTSPRIQPDYPAERSSLIPISGHRASPNPVAMET 473
               ++   ++  +TT+   P        +       T  SP       P  P        ++   P S         SP P   M T
Sbjct:  1676 PPTTTPSSPITTTPSPPTTTM----TTPSPTTTPSSPITTTTT-PSSTTTPSPPPTTMTT 1730

Query:   474 RSDNR-PSVPVQFQYFLPTYPPSAYPLAAHTYTPITSSVSTIRQYPVSAQAPNS 526
              S       PS P             LP         S+ PL        T TP+    S++             P S     P +
Sbjct:  1731 PSPTTTPSPPTTTMTTLPPTTTSS-PL---TTTPLPPSITPPTFSPFSTTTPTT 1780

 Score = 189 (28.4 bits), Expect = 8.0e-09, P = 8.0e-09
 Identities = 92/374 (24%), Positives = 133/374 (35%)

Query:   439 THTSPRIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYF-LPTYPPSAY 497
             T + P    P    P      ++  +P  +        + P                         PS P+              LPT   PS
Sbjct:  1398 TPSPPTTTPSPPPTTTTTLPPTTTPSPPTTTTTTPPPTTTPSPPITTTTTPLPTTTPSP- 1456

Query:   498 PLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVASTVHLNPMQL-MTVDASHAR 556
             P++    T   P T++ S          P                  S   T   T          +T       PM    +T   AS
Sbjct:  1457 PISTTTTPPPTTTPSPPTTTPSPPTTTPSPPTTTTTTPPPTTTPSPPMTTPITPPASTTT 1516

Query:   557 HIQGIQPAPISTQGIQPAPIGTPGIQPAPLGTQGIHSATPINTQGLQPAPMGTQQPQPEG 616
                      P+P +T       P P   TP       +P   T    I        P +T   L P         T   P P
Sbjct:  1517 LPPTTTPSPPTTTTTTPPPTTTP----SPPTTTPI--TPPTSTTTLPP----TTTPSPPP 1566

Query:   617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAP--AQGSSPRPSILRKK 674
             T+              T     +P     P  +     P+      T+     T                 +T  +P          ++P  P+
Sbjct:  1567 TTTTT---PPPTTTPSP---PTTTTPSPPTITTTTPPPTTTPSPPTTTTTTPPPTTTPSP 1620

Query:   675 PATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAV-PPTAQQPPPTIPTMIAA--A 731
             P T       P +       +      P  T               +           PT          PPT       P P I T
Sbjct:  1621 PTTTPITPPTS--TTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPPITTTTTPPPT 1678

Query:   732 SPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPV------PEIKVK 785
             + PS P     + P         TP    TT     ++P    + T          S ++             PP                        P
Sbjct:  1679 TTPSSPITTTPSPPTTTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPTTMTTPSPTTTPS 1738
```

```
Query:   786 EEVEPMDIMRPVSAVPPLATNTVSPSL 812
                 M  + P +   PL T  + PS+
Sbjct:  1739 PPTTTMTTLPPTTTSSPLTTTPLPPSI 1765

 Score = 185 (27.8 bits), Expect = 1.6e-09, P = 1.6e-09
 Identities = 71/270 (26%), Positives = 99/270 (36%)

Query:   563 PAPISTQGIQPAPIGTPGIQPAPLGTQGIHSATP---INTQGLQPAPMGTQQPQ---PEG 616
             P+P +T     P P  TP    P     T  + + TP    I+T    P P  T  P    P
Sbjct:  1422 PSPPTTTTTTPPPTTTPS-PPITTTTTPLPTTTPSPPISTT-TTPPPTTTPSPPTTTPSP 1479

Query:   617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSSPRPSILRKKPA 676
              T+          T     P + P   +P  TT + T    S +T  P    SP  +     P
Sbjct:  1480 PTTTPSPPTTTTTTPPPTTTP---SPPMTTPI-TPPASTTTLPPTTTPSPPTTTTTTPPP 1535

Query:   677 TDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVPPTAQQPPPTIPTMIAAASPPSQ 736
             T    P +       TP+T     T +      P+    P T    PPPT          + PS
Sbjct:  1536 TTTPSPPT------TTPITPPTSTTTLPPTTTPS-PPPTTTTTPPPTTTPSPPTTTTPSP 1588

Query:   737 PAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEVEPMDIMRP 796
             P +   +T P       +PP TT      PPP+ T        ++ + PP       +       P       P
Sbjct:  1589 PTITTTTPPPTTTPSPPTTT-TTTPPPTTTPSPPTTTPITPPTSTTTLPPTTTPSP--PP 1645

Query:   797 VSAVPPLATNTVSPSLALLANNLSMPTSDLPPGASP 832
              +    P  T T SP         +        T+   PP   +P
Sbjct:  1646 TTTTTPPPTTTPSPPTTTTPSPPITTTTTPPPTTTP 1681

 Score = 183 (27.5 bits), Expect = 3.4e-09, P = 3.4e-09
 Identities = 91/390 (23%), Positives = 139/390 (35%)

Query:   326 PSAAISIQRPAQSRDVTTR-ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPIL 384
             PS    +    P   +     T       T PS P     T            T        I +T TP+    T +P +
Sbjct:  1399 PSPPTTTPSPPPTTTTTLPPTTTPSPPTTTTTTPPPTTTPSPPITTTTTPLPTTTPSPPI 1458

Query:   385 ATNTIPSATTAGSVSHTQAPTSTIVTMTVPSHSSHATAVTTSNIP--VAKVVPQQITHTS 442
             +T T P  TT  S   T  P+    T + P+ ++   TT+  P      + P   T T
Sbjct:  1459 STTTTPPPTTTPSPP-TTTPSPPTTTPSPPTTTTTPPPTTTPSPPMTTPITPPASTTTL 1517

Query:   443 PRIQPDYPAERSSLIPISGHRASP---NPVAMETRSDNRP--SVPVQFQYFLPTYPPSAY 497
             P     P   ++  P     SP     P+    T +    P   + P         T PP+
Sbjct:  1518 PPTTTPSPPTTTTTTPPPTTTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTTTPPPTT 1577

Query:   498 PLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVASTVHLNPMQL-MTVDASHAR 556
             P    T TP    ++++T    P +  +  +P      T  T      +T     P     +T    S
Sbjct:  1578 PSPPTTTTPSPPTITTTTPPPTTTPSPP---TTTTTTPPPTTTPSPPTTTPITPPTSTTT 1634

Query:   557 HIQGIQPAPISTQGIQPAPIGTPGIQPAPLGTQGIHSATPINTQGLQPAPMGTQQPQPEG 616
                 P+P  T       P P  TP    P P  T         T    T          P           P
Sbjct:  1635 LPPTTTPSPPPTTTTTPPPTTTPS--P-PTTTTPSPPITTTTTPPPTTTPSSPITTTPSP 1691

Query:   617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSSPRPSILRKKPA 676
             T+ +      T   ++PI+    +    P++TT           + +T +P     SP  + +     P
Sbjct:  1692 PTTTMTTPSPTTTPSSPITT--TTTPSSTTTPSPPPTTMTTPSPTTTPSPPTTTMTTLPP 1749

Query:   677 TDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVPP 715
             T  + P +     +   P +++   T S   +    PT     P
Sbjct:  1750 TTTSSPLT----TTPLPPSITPPTFSPFSTTTPPTTPCVP 1784

 Score = 176 (26.4 bits), Expect = 1.8e-07, P = 1.8e-07
 Identities = 101/402 (25%), Positives = 142/402 (35%)

Query:   345 ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPILATNTIPSATTAGSVSHTQAP 404
             IT PS P    TP       T           +T +P      T   P     T     P    TT  +    T  P
Sbjct:  1396 ITTPSPPTT-TPSPPPTTTTTLPPTTTPSPPTTTTTTPPPTTTPSPPITTTTTPLPTTTP 1454

Query:   405 TSTIVTMTVPSHSSHATAVTTS-NIPVAKVVPQQITHTSPRIQPDYPAERSSLIPISGHR 463
             +  I T T P  ++  +   TT+ + P        P   T T+P    P                  PI+
Sbjct:  1455 SPPISTTTTPPPTTTPSPPTTTPSPPTTTPSPPTTTTTP--PPTTTPSPPMTTPITPP- 1511

Query:   464 ASPNPVAMETRSDNRPSVPVQFQYFLPTYPPSAYPLAAHTYTPITSSVSTIRQYPVSAQA 523
             AS    +   T       PS P        T PP+  P +  T TPIT    ST       P +   +
Sbjct:  1512 ASTTTLPPTTT----PSPPTTTT---TTPPPTTTP-SPPTTTPITPPTSTTTLPPTTTPS 1563

Query:   524 PNSAITAQ----TGVGVASTVHLNPMQLMTVDASHARHIQGIQPAPISTQGIQPAPIGTP 579
             P     T         T        +T     +P   + T                P+P +T      P P  TP
Sbjct:  1564 PPPTTTTTPPPTTTPSPPTTTTPSPPTITTTTPPPTT-----TPSPPTTTTTTPPPTTTP 1618

Query:   580 G------IQPAPLGTQGIHSAT---PINTQGLQPAPMGTQQPQPEGKTSAVVLADGATIV 630
                     I P P  T  +   T   P T      P P  T  P          S  +
Sbjct:  1619 SPPTTTPITP-PTSTTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPPITTTTTPPP 1677
```

```
Query:   631 ANPISNPFSAAPAA-TTVVQTHSQSASTNAP-AQGSSPRPSILRKKPATDGAKPKSEIHV 688
                 S+P +  P+   TT + T S + + ++P    ++P  +      P T    P
Sbjct:  1678 TTTPSSPITTTPSPPTTTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPTTMTTPSP---T 1734

Query:   689 SMATPVTVSMETVSNQNNDQPTIAVPPTAQQPPPTIPTMIAAASPPSQPAVALSTIPG 746
             +   +P T +M T+            P     P       PPT        + + P+ P V L     G
Sbjct:  1735 TTPSPPTTTMTTLPPTTTSSPLTTTPLPPSITPPTFSPF--STTTPTTPCVPLCNWTG 1790

 Score = 168 (25.2 bits), Expect = 9.3e-08, P = 9.3e-08
 Identities = 89/387 (22%), Positives = 133/387 (34%)

Query:   448 DYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFLPTYPPSAYPLAAHTYTPI 507
             DY      +   P+      +P+P    T + + P  P       PT  PS  P    T  P
Sbjct:  1381 DYKIRVNCCWPMDKCITTPSP---PTTTPSPP--PTTTTTLPPTTTPSP-PTTTTTTPPP 1434

Query:   508 TSSVS---TIRQYPVSAQAPNSAITAQTGVGVASTVHLNPMQLMTVDASHARHIQGIQPA 564
             T++ S    T      P+     P+  I+  T       +T      P      T   +          P+
Sbjct:  1435 TTTPSPPITTTTTPLPTTTPSPPISTTTTPPPTTT----PSPPTTTPSPPTT-----TPS 1485

Query:   565 PISTQGIQPAPIGTPGI-QPAPLGTQGIHSATPINTQGLQPAPMGTQQPQ---PEGKTSA 620
             P +T       P  P    TP          P+            +    P     T         P          T       P       P      T+
Sbjct:  1486 PPTTTTTTPPPTTTPSPPMTTPITPPASTTTLPPTTTPSPPTTTTTTPPPTTTPSPPTTT 1545

Query:   621 VVLADGATIVANPISNPFSAAPAATTVVQTHSQSA-STNAPAQGS----SPRPSILRKKP 675
              +     +T     P  + P         TT       T + S    +T    P+   +      +P  P+          P
Sbjct:  1546 PITPPTSTTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPPTITTTTPPPTTTPSPP 1605

Query:   676 ATDGAKPKSEIHVS--MATPVTVSMETVSNQNNDQPTIAVPPTAQQPPPTIPTMIAAASP 733
             T       P         S        TP+T       T +            P+       P  T         PPPT                 +
Sbjct:  1606 TTTTTPPPTTTPSPPTTTPITPPTSTTTLPPTTTPS-PPPTTTTPPPTTTPSPPTTTT 1664

Query:   734 PSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGP----PVPEIKVKEEVE 789
             PS P      +T P         + PITT    + P     ++T           ++       P            P
Sbjct:  1665 PSPPITTTTTPPPTTTPSSPITTTPSPPTTMTTPSPTTTPSSPITTTTTPSSTTTPSPP 1724

Query:   790 PMDIMRPVSAVPPLATNTVSPSLALLANNLSMPTSDLPPGASP 832
             P   +   P        P        T      +L          +   +  T+   LPP    +P
Sbjct:  1725 PTTMTTPSPTTTPSPPTTMTTLPPTTTSSPLTTTPLPPSITP 1767

 Score = 154 (23.1 bits), Expect = 2.7e-06, P = 2.7e-06
 Identities = 70/277 (25%), Positives = 92/277 (33%)

Query:   565 PISTQGIQPAPIGTPGIQPAPLGTQGIHSATPINTQGLQPAPMGTQQPQPEGKTSAVVLA 624
             PIST      P P   TP    P P   T       + TP                P  T  P P    T     +
Sbjct:  1457 PISTT-TTPPPTTTPS--P-PTTTPSPPTTTPSPPTTTTTPPPTTTPSPPMTTP--ITP 1510

Query:   625 DGATIVANPISNPFSAAPAATTVVQTHSQSASTNAP----AQGSSPRPSILRKKPATDGA 680
             +T     P + P          TT      T + S   T   P                 ++    P+         P T
Sbjct:  1511 PASTTTLPPTTTPSPPTTTTTTPPPTTTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTT 1570

Query:   681 KPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVPPTAQQ--PPPTIPTMIAAASPPSQPA 738
               P          S   T   T   S   T++ +              T     PPT              PPPT   T           + P    P
Sbjct:  1571 TPPPTTTPSPPTTTTPSPPTITTTTPPPTTTPSPPTTTTTTPPPTT-TPSPPTTTPITPP 1629

Query:   739 VALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEVEPMDIMRPVS 798
              +   +T+P          +PP    TT           PPP+    T              ++          PP+                 +
Sbjct:  1630 TSTTTLPPTTTPSPPPTT-TTTPPPTTTPSPPTTTPSPPITTTTTPPPTTTPSSPITTT 1688

Query:   799 AVPPLATNTV-----SPSLALLANNL--SMPTSDLPPGASPRKKP 836
               PP    T    T               +PS    +             S    T        PP             P
Sbjct:  1689 PSPPTTTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPTTMTTPSP 1733

 Score = 148 (22.2 bits), Expect = 1.1e-05, P = 1.1e-05
 Identities = 62/254 (24%), Positives = 89/254 (35%)

Query:   583 PAPLGTQGIHSATPINTQGLQPAPMGTQQPQPEGKTSAV-----VLADGATIVANPISNP 637
             P+P  T      S P  T   L P       T P P     T+             +   T        P+
Sbjct:  1399 PSPPTTTP--SPPPTTTTTLPP----TTTPSPPTTTTTPPPTTTPSPPITTTTTPLPTT 1452

Query:   638 FSAAPAATTVVQTHSQSASTNAPAQGSSPRPSILRKKPATDGAKPKSEIHVS--MATPVT 695
               + P +TT      T        + + + P         SP P+          P T         P                S    M  TP+T
Sbjct:  1453 TPSPPISTTT--TPPPTTTPSPPTTTPSP-PTTTPSPPTTTTTPPPTTTPSPPMTTPIT 1509

Query:   696 VSMETVSNQNNDQPTIAVPPTAQQPPPTIPTMIAAASPPSQPAVALSTIPGAVPITPPIT 755
                  T +          P+        T        PP  T  P+          + P    P     + +T+P            +PP T
Sbjct:  1510 PPASTTTLPPTTTPSPPTTTTTPPPTTTPS--PPTTPITPPTSTTTLPPTTTPSPPT 1567

Query:   756 TIAAAPPPSVTVGGSLSSVLGPPVPEIKVKEEVEPMDIMRPVSAVPPLATNTVSPSLALL 815
             T        PPP+  T           ++           PP                                       +    PP     T          P+     +
Sbjct:  1568 T-TTTPPPTTTPSPPTTTTPSPPTITTTTPPPTTTPSPPTTTTTTPPPTTTPSPPTTTPI 1626
```

```
Query:   816 ANNLSMPTSDLPPGASPRKKP 836
              S  T+ LPP  +P   P
Sbjct:  1627 TPPTS--TTTLPPTTTPSPPP 1645

 Score = 131 (19.7 bits), Expect = 1.2e-03, P = 1.2e-03
 Identities = 112/492 (22%), Positives = 174/492 (35%)

Query:    96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154
             V   P P       T    + + T V  T  P    TP       +    + P P PT    P
Sbjct:  3977 VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT 4036

Query:   155 A-PAPPSTLSLPPKVP-GQVTVTMESSIPQASAIPVATISGQQGHPSNLHHIMTTNVQMS 212
               P   +T    P   P G  T T           +  P  T +G Q  P+       TT V  +
Sbjct:  4037 TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-PTTTPITTTTTVTPT 4095

Query:   213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268
                +     P    P+     +   P                +++ +TT     T       T  P    I
Sbjct:  4096 PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 4155

Query:   269 IHQPIQSRPPVTTSNAIPPA--VVATVSATRAQSPVITTTA--AHATDSALSRPTLSIQH 324
                   +  P  T      P      +  T + T      +P  T T     H + + ++  T S
Sbjct:  4156 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTGPPTHTSTAPIAELTTSNPP 4215

Query:   325 PPSAAISIQRPAQS--RDVTTRI-TLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVA 381
             P S+       R    S    + TT + TLP  PA+     +    T       T   +   T        T++
Sbjct:  4216 PESSTPQTSRSTSSPLTESTTLLSTLP--PAI----EMTSTAPPSTPTAPTTTSGGHTLS 4269

Query:   382 PILATNTIPSAT-TAGSVS-HTQAPTSTIVTMTVPSHSSHATAVTTSNIPVAKVVPQQIT 439
             P  +T T  P    T  T G+ +  + APT + V   T   S       A T +  P++     P  I
Sbjct:  4270 PPPSTTTSPPGTPTRGTTTGSSSAPTPSTVQTTTTS------AWTPTPTPLS--TPSIIR 4321

Query:   440 HTSPRIQPDYPAERSSLIPISGHRASPNP-VAMETRSDN----RPSVPVQFQYFLPTYP- 493
              T     ++P  YP+          ++         +P     V     T  D           S+       +++    + P
Sbjct:  4322 TTG--LRP-YPSSVLICCVLNDTYYAPGEEVYNGTYGDTCYFVNCSLSCTLEFYNWSCPS 4378

Query:   494 -PSAYPLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVASTVHLNPMQLMTVDA 552
              PS   P   +  +  TP   S   S+      P        P                  T   L     +  T
Sbjct:  4379 TPSPTPTPSKS-TPTPSKPSSTPSKPTPGTKPPECPDFDPPRQENETWWLCDCFMATCKY 4437

Query:   553 SHARHIQGIQ----PAPISTQGIQPAPIGTP 579
             ++    I    ++          P  P    + G+QP   +   P
Sbjct:  4438 NNTVEIVKVECEPPPMPTCSNGLQPVRVEDP 4468

 Score = 117 (17.6 bits), Expect = 1.8e-02, P = 1.8e-02
 Identities = 41/156 (26%), Positives = 55/156 (35%)

Query:   710 TIAVPPTAQQPPPTIPTMIAAASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGG 769
             T + P T     PPPT   T +     + PS P      +T P       +PPITT       P P+ T
Sbjct:  1398 TPSPPTTTPSPPPTTTTLPPTTTPSPPTTTTTPPPTTTPSPPITT-TTTPLPTTTPSP 1456

Query:   770 SLSSVLGPPVPEIKVKEEVEPMDIMRPVSAVPPLATNTVSPSLALLANNLSMPTSDLPPG 829
             +S+      PP               P       P  +      P          T  T  SP                T+   PP
Sbjct:  1457 PISTTTTPP-----------PTTTPSPPTTTPSPPTTTPSPPTTTTTP-PPTTTPSPPM 1504

Query:   830 ASPRKKPRKQQHVISTEEGDMMETNSTDDEKSTAKS 865
             +P      P       +    T            T  +T     +T   S
Sbjct:  1505 TTPITPPASTTTLPPTTTPSPPTTTTTTPPPTTTPS 1540

 Score = 61 (9.2 bits), Expect = 1.6e-09, P = 1.6e-09
 Identities = 23/93 (24%), Positives = 41/93 (44%)

Query:   397 SVSHTQAPTSTIVTMTVPSHSSHATAVTTSNIPVAKVV----PQQITHTSPRIQPDYPAE 452
             S++    +   +T   T+T+P++    +    T     TT+   P   +   V         P+     S    I    D+P+
Sbjct:  1257 SITTRPSTLTTFTTITLPTTPTSFTTTTTTTTPTSSTVLSTTPKLCCLWSDWINEDHPSS 1316

Query:   453 RSS---LIPISGHRASPNPVAMETRSDNRPSVPVQ 484
             S         P  G    +P    +    E RS     P  +   ++
Sbjct:  1317 GSDDGDREPFDGVCGAPEDI--ECRSVKDPHLSLE 1349

 Score = 50 (7.5 bits), Expect = 8.0e-09, P = 8.0e-09
 Identities = 16/41 (39%), Positives = 19/41 (46%)

Query:   334 RPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKTIFSTGTP 374
             RP+       TT ITLP+ P     T          T     T+ ST TP
Sbjct:  1261 RPSTLTTFTT-ITLPTTPTSFTTTTTTTTPTSSTVLST-TP 1299

 Score = 46 (6.9 bits), Expect = 5.4e-08, P = 5.4e-08
 Identities = 24/106 (22%), Positives = 37/106 (34%)

Query:   324 HPPSAAISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPI 383
             +PP A++  +    +S    T      +     P   G      Q         A          G           I
```

```
Sbjct:  1196 YPPGASVPTEETCKSCVCTNSSQVVCRPEEGKILNQTQDGAFCYWEICGPNGTVEKHFNI 1255

Query:   384 LATNTIPSA-TTAGSVSHTQAPTSTIVTMTVPSHSSHATAVTTSNI 428
             +  T PS  TT  +++     PTS   T T   + +S     TT  +
Sbjct:  1256 CSITTRPSTLTTFTTITLPTTPTSFTTTTTTTTPTSSTVLSTTPKL 1301

 Score = 44 (6.6 bits), Expect = 8.7e-08, P = 8.7e-08
 Identities = 14/34 (41%), Positives = 17/34 (50%)

Query:   478 RPSVPVQFQYF-LPTYPPSAYPLAAHTYTPITSSV 511
             RPS     F    LPT P S +      T TP +S+V
Sbjct:  1261 RPSTLTTFTTITLPTTPTS-FTTTTTTTTPTSSTV 1294
```

Pedant information for DKFZphtes3_2a11, frame 2

Report for DKFZphtes3_2a11.2

```
[LENGTH]        1048
[MW]            110324.04
[pI]            9.83
[HOMOL]         PIR:I47141 gastric mucin (clone PGM-2A) - pig (fragment) 8e-15
[FUNCAT]        30.90 extracellular/secretion proteins      [S. cerevisiae, YIR019c] 1e-09
[FUNCAT]        30.01 organization of cell wall      [S. cerevisiae, YIR019c] 1e-09
[FUNCAT]        01.05.01 carbohydrate utilization      [S. cerevisiae, YIR019c] 1e-09
[FUNCAT]        30.02 organization of plasma membrane      [S. cerevisiae, YDR420w] 4e-09
[FUNCAT]        01.05.04 regulation of carbohydrate utilization      [S. cerevisiae, YDR420w]
4e-09
[FUNCAT]        98 classification not yet clear-cut   [S. cerevisiae, YJR151c] 4e-06
[FUNCAT]        03.04 budding, cell polarity and filament formation  [S. cerevisiae, YGR014w]
1e-05
[FUNCAT]        11.01 stress response  [S. cerevisiae, YHL028w] 1e-04
[FUNCAT]        09.01 biogenesis of cell wall      [S. cerevisiae, YHL028w] 1e-04
[EC]            3.2.1.3 Glucan 1,4-alpha-glucosidase 3e-08
[PIRKW]         glycosidase 3e-08
[PIRKW]         transmembrane protein 3e-08
[PIRKW]         polysaccharide degradation 3e-08
[PIRKW]         glycoprotein 9e-08
[PIRKW]         calcium binding 9e-08
[PIRKW]         hydrolase 3e-08
[PIRKW]         cytoskeleton 7e-08
[SUPFAM]        equine herpesvirus glycoprotein X 2e-07
[SUPFAM]        yeast glucan 1,4-alpha-glucosidase homolog 3e-08
[SUPFAM]        polymorphic epithelial mucin 7e-08
[SUPFAM]        glucan 1,4-alpha-glucosidase homology 3e-08
[SUPFAM]        equine herpesvirus 1 glycoprotein homology 2e-07
[PROSITE]       MYRISTYL        9
[PROSITE]       AMIDATION       1
[PROSITE]       CAMP_PHOSPHO_SITE       2
[PROSITE]       CK2_PHOSPHO_SITE        10
[PROSITE]       PKC_PHOSPHO_SITE        12
[PROSITE]       ASN_GLYCOSYLATION       3
[KW]            Irregular
[KW]            LOW_COMPLEXITY    20.04 %
```

```
SEQ     MGPPRHPQAGEIEAGGAGGGRRLQVEMSSQQFPRLGAPSTGLSQAPSQIANSGSAGLINP
SEG     ........xxxxxxxxxxxx........................................
PRD     cccccccccccccccccccceeeeeecccccccccccccccccccccccccccccccccc

SEQ     AATVNDESGRDSEVSAREHMSSSSSLQSREEKQEPVVVRPYPQVQMLSTHHAVASATPVA
SEG     ....................xxxxx.......................xxxxxxxxxxxx
PRD     cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ     VTAPPAHLTPAVPLSFSEGLMKPPPKPTMPSRPIAPAPPSTLSLPPKVPGQVTVTMESSI
SEG     xxxxxxxxxxxx.......xxxxxxxxxx..xxxxxxxxxxxxxx..............
PRD     cccccccccccccccccccccccccccccccccccccccccccccccccceeecccccc

SEQ     PQASAIPVATISGQQGHPSNLHHIMTTNVQMSIIRSNAPGPPLHIGASHLPRGAAAAAVM
SEG     .....................................................xxxxx..
PRD     cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ     SSSKVTTTVLRPTSQLPNAATAQPAVQHIIHQPIQSRPPVTTSNAIPPAVVATVSATRAQS
SEG     ............................................................
PRD     cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ     PVITTTAAHATDSALSRPTLSIQHPPSAAISIQRPAQSRDVTTRITLPSHPALGTPKQQL
SEG     ............................................................
PRD     cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ  HTMAQKTIFSTGTPVAAATVAPILATNTIPSATTAGSVSHTQAPTSTIVTMTVPSHSSHA
SEG  ............xxxxxxxxxx................................xxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  TAVTTSNIPVAKVVPQQITHTSPRIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPS
SEG  xxxxxx......................................................
PRD  cccccccccccccccccccccccccccccccccccccccccccceeeeccccccccc

SEQ  VPVQFQYFLPTYPPSAYPLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVASTV
SEG  ............................................................
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  HLNPMQLMTVDASHARHIQGIQPAPISTQGIQPAPIGTPGIQPAPLGTQGIHSATPINTQ
SEG  ............................................................
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  GLQPAPMGTQQPQPEGKTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAP
SEG  ............................................................
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  AQGSSPRPSILRKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNNDQPTIAVPPTAQQP
SEG  ...............................................xxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  PPTIPTMIAAASPPSQPAVALSTIPGAVPITPPITTIAAAPPPSVTVGGSLSSVLGPPVP
SEG  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccceeeccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  EIKVKEEVEPMDIMRPVSAVPPLATNTVSPSLALLANNLSMPTSDLPPGASPRKKPRKQQ
SEG  xxxxxxxxxx....................xxxxxxxxxx....................
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  HVISTEEGDMMETNSTDDEKSTAKSLLVKAEKRKSPPKEYIDEEGVRYVPVRPRPPITLL
SEG  .............................................xxxxxxxxxxx....
PRD  ccccccccccccccccccccchhhhhhhhcccccccccccccccccccccccccccccee

SEQ  RHYRNPWKAAYHHFQRYSDVRVKEEKKAMLQEIANQKGVSCRAQGWKVHLCAAQLLQLTN
SEG  ............................................................
PRD  eecccccchhhhhhccccchhhhhhhhhhhhhhhhhhhccceeecccceeehhhhhhhhhc

SEQ  LEHDVYERLTNLQEGIIPKKKAATDDDLHRINELIQGNMQRCKLVMDQISEARDSMLKVL
SEG  ............................................................
PRD  cchhhhhhhhhhhceeeecccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  DHKDRVLKLLNKNGTVKKVSKLKRKEKV
SEG  ...............xxxxxxxxxxxxx
PRD  hhhhhhhhhccccceeeeeeeeeccccc
```

Prosite for DKFZphtes3_2al1.2

PS00001	818->822	ASN_GLYCOSYLATION	PDOC00001
PS00001	854->858	ASN_GLYCOSYLATION	PDOC00001
PS00001	1033->1037	ASN_GLYCOSYLATION	PDOC00001
PS00004	872->876	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1037->1041	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	68->71	PKC_PHOSPHO_SITE	PDOC00005
PS00005	75->78	PKC_PHOSPHO_SITE	PDOC00005
PS00005	242->245	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	355->358	PKC_PHOSPHO_SITE	PDOC00005
PS00005	442->445	PKC_PHOSPHO_SITE	PDOC00005
PS00005	513->516	PKC_PHOSPHO_SITE	PDOC00005
PS00005	665->668	PKC_PHOSPHO_SITE	PDOC00005
PS00005	831->834	PKC_PHOSPHO_SITE	PDOC00005
PS00005	862->865	PKC_PHOSPHO_SITE	PDOC00005
PS00005	940->943	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1035->1038	PKC_PHOSPHO_SITE	PDOC00005
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	75->79	CK2_PHOSPHO_SITE	PDOC00006
PS00006	88->92	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	473->477	CK2_PHOSPHO_SITE	PDOC00006
PS00006	844->848	CK2_PHOSPHO_SITE	PDOC00006
PS00006	855->859	CK2_PHOSPHO_SITE	PDOC00006
PS00006	959->963	CK2_PHOSPHO_SITE	PDOC00006
PS00006	984->988	CK2_PHOSPHO_SITE	PDOC00006
PS00008	15->21	MYRISTYL	PDOC00008

PS00008	16->22	MYRISTYL	PDOC00008
PS00008	36->42	MYRISTYL	PDOC00008
PS00008	233->239	MYRISTYL	PDOC00008
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	535->541	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00008	768->774	MYRISTYL	PDOC00008
PS00009	19->23	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_2al1.2)

DKFZphtes3_2a17
----

group: metabolism

DKFZphtes3_2a17 encodes a novel 574 amino acid protein without similarity to known proteins.

The novel protein contains a thiol protease cys pattern. Eukaryotic thiol proteases (EC 3.4.22.-) are a family of proteolytic enzymes containing an active site cysteine. Cathepsins belong to this protease family.

The new protein can find application in modulation of proteolytic processes and as a new enzyme for proteomic analysis and biotechnologic production processes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 2312 bp
Poly A stretch at pos. 2300, polyadenylation signal at pos. 2273

```
   1 GTTTTCACCT GATCATTAGA AACTAATGAA ACACCTTTTA AGTCTTATGA
  51 ATTCAGGTTA CACTGTTTTC CAGATGCCTT GGCAGCTGGT ACAGGGCCTC
 101 TGAAAAATGG AACCAAATTC TCTGAGGACT AAAGTCCCAG CTTTCTTATC
 151 TGATTTGGGG AAGGCCACAT TGAGGGGAAT CAGAAAGTGT CCCCGATGTG
 201 GCACATACAA TGGAACCCGG GGACTGAGCT GTAAGAACAA GACATGTGGA
 251 ACCATATTCC GCTACGGTGC ACGCAAGCAG CCTAGTGTTG AAGCTGTCAA
 301 AATCATTACA GGCTCTGATC TTCAGGTCTA CTCAGTGCGG CAAAGAGACC
 351 GGGGCCCTGA TTACCGATGC TTTGTGGAGC TCGGGGTTTC AGAGACAACA
 401 ATCCAGACAG TGGATGGGAC GATCATCACT CAGCTGAGCT CTGGACGGTG
 451 TTATGTCCCC TCATGCCTGA AAGCTGCCAC TCAAGGCGTT GTGGAAAACC
 501 AGTGCCAGCA CATCAAGCTG GCGGTGAACT GCCAGGCAGA GGCCACCCCT
 551 CTGACCCTGA AGAGCTCGGT CCTGAATGCA ATGCAGGCCT CCCCGGAAAC
 601 CAAACAGACC ATCTGGCAGT TGGCCACGGA ACCCACAGGT CCTCTGGTGC
 651 AGAGAATTAC TAAAAACATC TTGGTGGTGA AATGCAAGGC AAGCCAGAAG
 701 CACAGTTTGG GGTATTTGCA TACATCTTTT GTGCAGAAAG TCAGTGGCAA
 751 AAGCTTGCCT GAGCGCCGCT TCTTCTGCTC CTGTCAGACT CTGAAATCGC
 801 ACAAGTCAAA TGCCTCCAAG GATGAGACAG CCCAGAGATG CATTCATTTC
 851 TTTGCTTGCA TCTGTGCCTT TGCCAGTGAT GAGACACTGG CTCAGGAATT
 901 CTCAGACTTC CTAAATTTTG ATTCCAGCGG TCTTAAAGAG ATTATTGTAC
 951 CCCAGTTAGG TTGCCATTCA GAATCAACAG TATCTGCTTG TGAGTCTACT
1001 GCCTCTAAGT CAAAGAAGAG GAGAAAGGAT GAAGTATCTG GTGCACAGAT
1051 GAACAGTTCA CTACTGCCTC AAGATGCAGT GAGCAGTAAT CTAAGGAAAA
1101 GTGGCCTGAA AAAGCCTGTG GTTGCTTCCT CGTTAAAAAG GCAGGCCTGT
1151 GGTCAGCTGT TAGATGAGGC ACAAGTGACT TTATCCTTCC AAGACTGGCT
1201 GGCCAGTGTC ACAGAACGCA TCCATCAAAC CATGCACTAT CAGTTTGATG
1251 GCAAACCAGA ACCATTGGTG TTCCACATTC CTCAGTCATT TTTTGATGCC
1301 CTGCAACAAA GAATATCTAT AGGAAGTGCA AAAAAACGGC TCCCCAACTC
1351 CACCACAGCT TTTGTTCGGA AAGATGCCTT GCCACTGGGA ACCTTTTCCA
1401 AGTATACTTG GCATATCACT AATATCCTGC AAGTTAAACA AATCTTAGAT
1451 ACCCCAGAGA TGCCCTTGGA AATCACCCGT AGCTTTATCC AGAACCGAGA
1501 TGGGACTTAT GAGCTATTTA AATGCCCTAA AGTGGAAGTA GAAAGCATAG
1551 CAGAAACCTA CGGTCGTATA GAAAAACAAC CAGTGCTGCG ACCCTTGGAA
1601 CTAAAAACTT TTCTCAAAGT TGGCAACACT TCCCCAGATC AAAAGGAGCC
1651 AACACCTTTC ATCATCGAGT GGATCCCAGA TATCCTTCCC CAATCTAAGA
1701 TTGGCGAGCT GCGGATCAAG TTTGAGTATG GCCACCACCG GAATGGGCAT
1751 GTGGCGGAGT ACCAAGACCA GCGGCCCCCC TTGGACCAGC CCTTGGAACT
1801 GGCCCCTCTG ACCACTATTA CTTTCCCTTA AAGCAAAACA AGATAATAAT
1851 CTTTTGCTGC TTAATTTGCA CATCCCCACC CCTTGACAAC TTTAAATGCT
1901 AGTTAGGCAC TTAGATGGCC CTGTTCCTTG GTAAACTGCT CTTAGCTAAG
1951 ATGCAAATTC TCAGTGCTTT CAAGTGGATT CTGTTGAAGA AAATCTCTTG
2001 TAAATAGCCT TTTTGATGCT GCTGTGTACA GTCTTCATTA TGCATTGGGC
2051 AGTATTTCTG GCTAGAGTTT TAAAAGGAAC AGAAAGAAAA CCAGCTTATT
2101 TTCCTTCTTA CGGACTCATC TTTAGCGTTT ATTTCAACCT TTTGCTAATT
2151 CTCTGAGAAA TCTGCAGCAC TCAGCCATAC ACCAACAGTG TTGGAAAGTT
2201 AACACCCTGG TTAGGGCAGA ATGTTAAAGA CCATCTTGGC AGAGTTCCAG
2251 CCACGCTCTT TATTCTGTTC TCAAATAAAG CAGTGTCACT AGTTTTTCCT
2301 AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 107 bp to 1828 bp; peptide length: 574
Category: putative protein

```
  1 MEPNSLRTKV PAFLSDLGKA TLRGIRKCPR CGTYNGTRGL SCKNKTCGTI
 51 FRYGARKQPS VEAVKIITGS DLQVYSVRQR DRGPDYRCFV ELGVSETTIQ
101 TVDGTIITQL SSGRCYVPSC LKAATQGVVE NQCQHIKLAV NCQAEATPLT
151 LKSSVLNAMQ ASPETKQTIW QLATEPTGPL VQRITKNILV VKCKASQKHS
201 LGYLHTSFVQ KVSGKSLPER RFFCSCQTLK SHKSNASKDE TAQRCIHFFA
251 CICAFASDET LAQEFSDFLN FDSSGLKEII VPQLGCHSES TVSACESTAS
301 KSKKRRKDEV SGAQMNSSLL PQDAVSSNLR KSGLKKPVVA SSLKRQACGQ
351 LLDEAQVTLS FQDWLASVTE RIHQTMHYQF DGKPEPLVFH IPQSFFDALQ
401 QRISIGSAKK RLPNSTTAFV RKDALPLGTF SKYTWHITNI LQVKQILDTP
451 EMPLEITRSF IQNRDGTYEL FKCPKVEVES IAETYGRIEK QPVLRPLELK
501 TFLKVGNTSP DQKEPTPFII EWIPDILPQS KIGELRIKFE YGHHRNGHVA
551 EYQDQRPPLD QPLELAPLTT ITFP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2a17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2a17, frame 2

Report for DKFZphtes3_2a17.2

```
[LENGTH]    574
[MW]        64076.89
[pI]        9.15
[PROSITE]   MYRISTYL            5
[PROSITE]   CK2_PHOSPHO_SITE    9
[PROSITE]   PKC_PHOSPHO_SITE    14
[PROSITE]   ASN_GLYCOSYLATION   5
[PROSITE]   THIOL_PROTEASE_CYS  1
[KW]        Alpha_Beta
```

```
SEQ  MEPNSLRTKVPAFLSDLGKATLRGIRKCPRCGTYNGTRGLSCKNKTCGTIFRYGARKQPS
PRD  cccccccccchhhhhccccchhhhhhccccccccccccccccccccccceeeecccccccc

SEQ  VEAVKIITGSDLQVYSVRQRDRGPDYRCFVELGVSETTIQTVDGTIITQLSSGRCYVPSC
PRD  ceeeeeeeccccceeeeeccccccccccceeeeeeccccccceeeccceeeeeeccccccchh

SEQ  LKAATQGVVENQCQHIKLAVNCQAEATPLTLKSSVLNAMQASPETKQTIWQLATEPTGPL
PRD  hhhhhhhhcchhhhheeehhhhhhhccccchhhhhhhhhccchhhhhhhhhcccccch

SEQ  VQRITKNILVVKCKASQKHSLGYLHTSFVQKVSGKSLPERRFFCSCQTLKSHKSNASKDE
PRD  hhhhhhheeeeeecccccccccccccceeeeeeecccccccceeeeccccccccccccccc

SEQ  TAQRCIHFFACICAFASDETLAQEFSDFLNFDSSGLKEIIVPQLGCHSESTVSACESTAS
PRD  hhhhhhhhhhhhhhhhhchhhhhhhhhhhcccccceeeeeeccccccccceeeccccccc

SEQ  KSKKRRKDEVSGAQMNSSLLPQDAVSSNLRKSGLKKPVVASSLKRQACGQLLDEAQVTLS
PRD  ccchhhhhcccccccccccccccchhhhhhhcccccceeehhhhhhhhhchhhhhhhhh

SEQ  FQDWLASVTERIHQTMHYQFDGKPEPLVFHIPQSFFDALQQRISIGSAKKRLPNSTTAFV
PRD  hhhhhhhhhhhhhhhhhhhccccccccceeehhhhhhhhhhhhhhhcccccccccccceeee

SEQ  RKDALPLGTFSKYTWHITNILQVKQILDTPEMPLEITRSFIQNRDGTYELFKCPKVEVES
PRD  ecccccccccceeeeehhhhhhhhhhccccccccceeeeeeccccceeeecccceeeeh

SEQ  IAETYGRIEKQPVLRPLELKTFLKVGNTSPDQKEPTPFIIEWIPDILPQSKIGELRIKFE
PRD  hhhhhhhhhcccccccccceeeeeccccccccccccceeeeeecccccccccccceeeeee
```

SEQ YGHHRNGHVAEYQDQRPPLDQPLELAPLTTITFP
PRD ecccccceeeeccccccccccccccccceeeccc

Prosite for DKFZphtes3_2a17.2

PS00001	35->39	ASN_GLYCOSYLATION	PDOC00001
PS00001	44->48	ASN_GLYCOSYLATION	PDOC00001
PS00001	235->239	ASN_GLYCOSYLATION	PDOC00001
PS00001	316->320	ASN_GLYCOSYLATION	PDOC00001
PS00001	414->418	ASN_GLYCOSYLATION	PDOC00001
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	41->44	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	150->153	PKC_PHOSPHO_SITE	PDOC00005
PS00005	196->199	PKC_PHOSPHO_SITE	PDOC00005
PS00005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS00005	228->231	PKC_PHOSPHO_SITE	PDOC00005
PS00005	231->234	PKC_PHOSPHO_SITE	PDOC00005
PS00005	302->305	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	369->372	PKC_PHOSPHO_SITE	PDOC00005
PS00005	407->410	PKC_PHOSPHO_SITE	PDOC00005
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	360->364	CK2_PHOSPHO_SITE	PDOC00006
PS00006	367->371	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	480->484	CK2_PHOSPHO_SITE	PDOC00006
PS00006	508->512	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	104->110	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008
PS00008	312->318	MYRISTYL	PDOC00008
PS00139	109->121	THIOL_PROTEASE_CYS	PDOC00126

(No Pfam data available for DKFZphtes3_2a17.2)

DKFZphtes3_2d15
----

group: testes derived

DKFZphtes3_2d15 encodes a novel 274 amino acid protein with similarity to C.elegans cosmid F25H2.1.

The novel protein contains a Pfam predicted C2-domain.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans F25H2.1

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3615 bp
Poly A stretch at pos. 3603, polyadenylation signal at pos. 3578

```
   1 GCGGCGGCCT CGAGGTGACA ACTGTCTCCG TCGCAGGCTC CGGCGGGGGC
  51 GCAGGAGGTC GCCCGGCGCG TCACTGTCGG GTCGGCGAGC CACGGGGGCC
 101 GCCGCAGCAC CATGGCGACC ACCGTCAGCA CTCAGCGCGG GCCGGTGTAC
 151 ATCGGTGAGC TCCCGCAGGA CTTCCTCCGC ATCACGCCCA CACAGCAGCA
 201 GCGGCAGGTC CAGCTGGACG CCCAGGCGGC CCAGCAGCTG CAGTACGGAG
 251 GCGCAGTGGG CACCGTGGGC CGACTGAACA TCACGGTGGT ACAGGCAAAG
 301 TTGGCCAAGA ATTACGGCAT GACCCGCATG GACCCCTACT GCCGACTGCG
 351 CCTGGGCTAC GCGGTGTACG AGACGCCCAC GGCACACAAT GGCGCCAAGA
 401 ATCCCCGCTG GAATAAGGTC ATCCACTGCA CGGTGCCCCC AGGCGTGGAC
 451 TCTTTCTATC TCGAGATCTT CGATGAGAGA GCCTTCTCCA TGGACGACCG
 501 CATTGCCTGG ACCCACATCA CCATCCCGGA GTCCCTGAGG CAGGGCAAGG
 551 TGGAGGACAA GTGGTACAGC CTGAGCGGGA GGCAGGGGGA CGACAAGGAG
 601 GGCATGATCA ACCTCGTCAT GTCCTACGCG CTGCTTCCAG CTGCCATGGT
 651 GATGCCACCC CAGCCCGTGG TCCTGATGCC AACAGTGTAC CAGCAGGGCG
 701 TTGGCTATGT GCCCATCACA GGGATGCCCG CTGTCTGTAG CCCCGGCATG
 751 GTGCCCGTGG CCCTGCCCCC GGCCGCCGTG AACGCCCAGC CCCGCTGTAG
 801 CGAGGAGGAC CTGAAAGCCA TCCAGGACAT GTTCCCCAAC ATGGACCAGG
 851 AGGTGATCCG CTCCGTGCTG GAAGCCCAGC GAGGGAACAA GGATGCCGCC
 901 ATCAACTCCC TGCTGCAGAT GGGGGAGGAG CCATAGAGCC TCTGCCTCGA
 951 TGCCGTTTTG CCCCCGCTCT TTGGACACGC CGACCCGGCG CTCCCCAAGG
1001 AATGCTGTCC CAACAAGATT CCCGTGAAAG AGCACCCGTG TCGCCCCCTC
1051 CCGTGGACTT CTGTGCCGCC CCGTCCACAC CTGTTCTTGG GTGCATGTGG
1101 GTTTTCGGTT CCTGGCGGTC CAGGACGGGG CGGGGGCTCC CCTCCCATCT
1151 CGTGCTGGGA GGTCTCAGCG CGCTCTCCTG TCCCTGGGAC GTGCGTCTCT
1201 CCTTCTCATG CCGTTCTGGA AAATGCTCTT GCTGTAGAGA GCAGCTGCTT
1251 CTGCCAGGGT GTTGGAGGTG GTGGAGCGCC TTCCGATTCC ATTCATGGCA
1301 TTTTGTGATG TGATGTAATT GGAATAGAGC TGTTGATTTA AGGCACACAC
1351 AATCCCTCAC ACTGTGGGTT TTTTTTAGAA CTTCCCAGAC GAAAACTCAC
1401 GCCCTTGCCC TAACGCGCTT TGCTGTGAGC CTGGCCCCTG CCCAGGGCTT
1451 GGGTCTGGTG AGCTGAGCAG CTTCCTGTGG ATGGTGTGGG GCCGGCCTCT
1501 GGCCTGGCTC ACCTGGCCAC TGTCCAGCCA GCCTTGTGAC AGACTCCGGC
1551 CTGAAGGCAG AATGAACCCA CACCTGGAGT GAGGAAGGGG GCCTGGCACG
1601 GTTGGCCAGG CTCTGCCTGA TTGCCAGCCA GCGGGCATCT GAAGCCGGGT
1651 CCTTCGCCCG CCGGAGGCTG CCGTCCGTCT CTCCTGCTGC GCTCGTGCCA
1701 GCTCCGTGGG TGTCCTCCCA GGGAGCTTCT CTTCTCAACA GGCCTTGCGA
1751 GGCTGGGGTG AGAGGTGATA GAGGCAGCAC TGTGCATGAT TCCGAGAGGG
1801 TGTGGTGGCA CTGCCAGCCG ACTGCTGACA GCTTGGGAGC TGCTGTGCCC
1851 AGGACGTGGG TTCAGCGTGG GCGAGGAAAG CCTGGCGAGC GTGGCCCTGT
1901 AAAAGCTTTC TGAGGCGGGA GGCGCTCACT TACCTCTGAC TGCCTGGGCG
1951 CTGCGTGTAG CATCTTGGCC TACAGGACAG ATTTTAGGTG ACACCTGGTT
2001 ATGACAGTCA GAAATTTGAG AAGCTTCTCA CAAGTGATGC ACTTTAAATA
2051 ATCTGCATGC CATTGAGACA CCTGCATGTC TGGTGTTTGT GGTTCAAGTG
2101 TCTTGCCGCC GGCCTTCGGA TGTAAACCCA CTGATAACGG ACAGAAAGAG
2151 AATGCCCACA AGTGGGTCTT CTGTGGAAGA TGCAGAAGGA GGAAGTTAGT
2201 GCTTACATTT TAGTCTTTTT CTCCCTCAAA AAAATAGGTT AAGTTTCAGT
2251 GCCAGCTAGA AAATACTGCT TTCTGCCATC GATTGGGGGT GGTTTTTGTC
2301 AAATATACTG TTGATAAATA TTTATTTTTG TAAACTTGAA GTGTGTGGTG
2351 GCCGTGGGGG AGGGACATGC TGGCAGCAGG CGCCTTCTTC AGCTGTGGGT
2401 CCTAAAGGCC TTTGATCCTT TGAAGAAGAA AGACATGGTA TTTGTTCAGC
2451 AGACGCCGAC CACTCAGACG GAGGGGCCCC TGGGATTCCC TGTCTCAGAT
2501 GGCCTGGTCT TACGCCTGTG TAGATTTCTT CTCCATTGGG AATGAAGGTG
2551 TCAGGCGGGA CTGGAACGTT CTAGATGGTA TGTTCCGTGA TATTAACAAC
2601 TCTAACCCAG GACAGACCAC AAGCCACACT CAGAGGCCTC ACTGTGCTGG
```

```
2651 GGGCTTCGGT GTCCAGGCGC CCAGGTGTGG CCACCAGCAC CGGTTTCTGC
2701 CTTCGCGTTG CTGGGGTGCA GTGAGACTGC CACACGCGTG CACATGTGGC
2751 TCTGTGGGTG TCTCCTAGAG AGGACGTGGC CCCTGCTGCC AGCCCTTGAG
2801 CAGCCCGTGT GGGGGCCCGA GGGACCCACA CAGTGGGGGC CAGCCTCGCT
2851 GGAGGGAGAG CAACCCTTTG CCGATGACCA CGCTTGCCGC CATCTCTTAG
2901 TTTTCTTTTT CACAAGCGCT TTATTTTTTT AATAGACAAA TCACATTTTG
2951 CAAGGCCTTT AATTAAATAA GATTCTTCTT TCCTTCATTT TATGCTTTAT
3001 TTCCTGTTTG AAGGCTTACT GTAGAAGTGG CTTACTGTAG AAGCAGCTTG
3051 CTGAGCCCCT CCGAGCGGTC CCCAGAATTA GCTGGTTCAC AACCCCCACC
3101 CTCCCCCGCC CCCGCCTGTG TCAGGTGTGG ATGAGGTCGT CACACTCAGA
3151 AGGACAGGCT TGTCTGCCAG CTCACAAGGG GAGGCTGCAG TGGGTTTGGG
3201 AGCTGGGTTT AGGCCCCTGG TGTCTGAGGG CCCAGGCCTT GCCAGCCTCT
3251 GCTGCTCCTG CTCCTGGGTT TGAAGATGCA GGCCGATCGC CAGCTCCGTG
3301 GCAGCGGTCA CTAAGGACAG CCTGACTGTG CCATCTTGGA GCCTCAGGCG
3351 GGGCTCCGGA GATAGAAGAC AGGTCGCCGG AGGCTCCCCC TCCTCTCCTC
3401 TCCCCTCTGC AGATGCTCCC TGGGCGCTAC CCTGCAGGGT GCCAGGCAGG
3451 AGTGGTCTCA GAACGTGCGC TTCTGATTAT TTTACTGGGG TCCATTGTCC
3501 AGATTTTTCT TTGATTGTAA AATATATTTT TACTTTTTAG TCTTCTAATT
3551 TAATAAATGA TCCATATAAA AATAGAGAAA TAAAGTCCTT TAAGGGAAGG
3601 TTTAAAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 112 bp to 933 bp; peptide length: 274
Category: similarity to unknown protein
Classification: no clue

```
  1 MATTVSTQRG PVYIGELPQD FLRITPTQQQ RQVQLDAQAA QQLQYGGAVG
 51 TVGRLNITVV QAKLAKNYGM TRMDPYCRLR LGYAVYETPT AHNGAKNPRW
101 NKVIHCTVPP GVDSFYLEIF DERAFSMDDR IAWTHITIPE SLRQGKVEDK
151 WYSLSGRQGD DKEGMINLVM SYALLPAAMV MPPQPVVLMP TVYQQGVGYV
201 PITGMPAVCS PGMVPVALPP AAVNAQPRCS EEDLKAIQDM FPNMDQEVIR
251 SVLEAQRGNK DAAINSLLQM GEEP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2d15, frame 1

TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2,
N = 1, Score = 385, P = 1.1e-35

>TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2
Length = 457

HSPs:

Score = 385 (57.8 bits), Expect = 1.1e-35, P = 1.1e-35
Identities = 77/182 (42%), Positives = 118/182 (64%)

```
Query:    4 TVSTQRGPVYIGELPQDFLRIT-PTQQQRQVQLDAQAAQQLQYGGAVGTVGRLNITVVQA 62
            TV+ +R  V +GELP  FLR+  P QQ  + ++  Q  + + +     T GRL++T+++A
Sbjct:    5 TVAERRRQVLVGELPPHFLRLAVPIQQTAEPEI-VQP-RMVSFVPP-NTRGRLSVTILEA 61

Query:   63 KLAKNYGMTRMDPYCRLRLGYAVYETPTAHNGAKNPRWNKVIHCTVPPGVDSFYLEIFDE 122
             L KNYG+ RMDPYCR+R+G   ++T  A N  + P WN+ ++  +P  V+S Y++IFDE
Sbjct:   62 NLVKNYGLVRMDPYCRVRVGNVEFDTNVAANAGRAPTWNRTLNAYLPMNVESIYIQIFDE 121

Query:  123 RAFSMDDRIAWTHITIPESLRQGKVEDKWYSLSGRQGDDKEGMINLVMSYAL--LPAAMV 180
            +AF  D+ IAW HI +P ++  G   D+++ LSG+QG+ KEGMI+L  S+A   LP
Sbjct:  122 KAFGPDEVIAWAHIMLPLAIFNGDNIDEYFQLSGQQGEGKEGMIHLHFSFAPIDLPLQQA 181
```

```
Query:   181 MPPQP 185
             P +P
Sbjct:   182 APAEP 186

 Score = 92 (13.8 bits), Expect = 1.8e-01, P = 1.7e-01
 Identities = 26/68 (38%), Positives = 38/68 (55%)

Query:   194 QQGVGYVPITGMPAVCSPGMVPV--ALP--PAAVNAQPRCSEEDLKAIQDMFPNMDQEVI 249
             QQG G    +  +     +P  +P+  A P  PA         +EED K IQ+MFP +D+EVI
Sbjct:   156 QQGEGKEGMIHLHFSFAPIDLPLQQAAPAEPAPAPLPVEITEEDTKEIQEMFPIVDKEVI 215

Query:   250 RSVLEAQR 257
             + +LE +R
Sbjct:   216 KCILEERR 223
```

Pedant information for DKFZphtes3_2d15, frame 1

Report for DKFZphtes3_2d15.1

```
[LENGTH]   274
[MW]       30281.97
[pI]       5.68
[HOMOL]    TREMBL:CEF25H2_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2 4e-36
[PFAM]     C2 domain
[KW]       Alpha_Beta
[KW]       LOW_COMPLEXITY    16.42 %

SEQ   MATTVSTQRGPVYIGELPQDFLRITPTQQQRQVQLDAQAAQQLQYGGAVGTVGRLNITVV
SEG   ..........................xxxxxxxxxxxxxxxxx.................
PRD   cccccccccceeeeecccccceeeeccccchhhhhhhhhhhhhhhhcccccceeeeceeeh

SEQ   QAKLAKNYGMTRMDPYCRLRLGYAVYETPTAHNGAKNPRWNKVIHCTVPPGVDSFYLEIF
SEG   ............................................................
PRD   hhhhhhhhcccccccchhhhheeeeeeccccccccccccccceeeeeccccccceeeeec

SEQ   DERAFSMDDRIAWTHITIPESLRQGKVEDKWYSLSGRQGDDKEGMINLVMSYALLPAAMV
SEG   ....................................................xxxxxxxx
PRD   ccccccccceeeeeccccccccccccccccceeeeecccccccccccceeeeehhhhhhhhc

SEQ   MPPQPVVLMPTVYQQGVGYVPITGMPAVCSPGMVPVALPPAAVNAQPRCSEEDLKAIQDM
SEG   xxxxxxxxxx......................xxxxxxxxxx.................
PRD   cccccceeeeeeeecccccccccccccceeeccccccccccccccceeeeccccchhhhhhhhhc

SEQ   FPNMDQEVIRSVLEAQRGNKDAAINSLLQMGEEP
SEG   ..................................
PRD   cccccchhhhhhhhhhcccccchhhhhhhhhhhccc
```

(No Prosite data available for DKFZphtes3_2d15.1)

Pfam for DKFZphtes3_2d15.1

```
HMM_NAME    C2 domain

HMM              *LtVrIIeARNLWkMDMnGfSDPYVKVdMdPdpkDtkKWKTkTiWNNGLN
                  L++++++A+  + + M+   DPY+++ +  + +    +T T +N   N
Query        55   LNITVVQAKLAKNYGMT-RMDPYCRLRLGYAVY-----ETPTAHNGAKN     97

HMM              PVWNEEeFvFedIPyPdlqrkMLRFaVWDWDRFSRBDFIGHCi*
                 P+WN    +     +P    + +     ++++D+  FS +D I+ +
Query        98  PRWN-KVIHCT-VPPGVDSF---YLEIFDERAFSMDDRIAWTH     135
```

DKFZphtes3_2e12
----

group: Transcription Factors

DKFZphtes3_2e12 encodes a novel 849 amino acid protein with similarity to Zinc finger proteins.

The new protein is a putative transcription factor with three C2H2 zinc fingers. Additionally, a cytochrome C family heme-binding site signature is present in the protein, which is only found in cytochrom C related proteins.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

similarity to finger proteins

complete cDNA, complete cds, 5 EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3205 bp
Poly A stretch at pos. 3192, polyadenylation signal at pos. 3171

```
   1 GGCACGGCCG GGTCCTGGCT GGCCAAACGA GGCTCGCGGA AGCAGCAGCC
  51 GCCGCCTGAC CGCAGCTGGA TTTTGAAGAT TGATCCAAGG GACTGTATTA
 101 ATTTCAGGAA TTGATTTGAA AGACACTGGC TCTGCCACTT AACAGCCATG
 151 TAACCTTGGA TATGGAAGAA AGTAGCAGTG TTGCCATGTT GGTGCCAGAT
 201 ATTGGGGAAC AGGAAGCTAT ACTGACTGCT GAAAGTATCA TCAGTCCTTC
 251 ATTGGAAATT GATGAACAAA GAAAAACTAA ACCAGATCCA TTAATCCATG
 301 TTATCCAGAA GTTAAGCAAG ATAGAAAAAT GAAAAGTCAC AAAAATGTCT
 351 TTTAATTGGG AAGAAACGCC CACGTTCAAG TGCTGCAACA CACTCTCTTG
 401 AAACCCAAGA ACTTTGTGAG ATTCCGGCTA AAGTAATCCA GTCACCTGCT
 451 GCTGATACTA GAAGGGCTGA GATGTCACAA ACAAATTTTA CCCCTGACAC
 501 TCTTGCCCAG AATGAAGGGA AGGCTATGTC TTATCAGTGT AGCCTTTGTA
 551 AGTTTCTATC ATCATCCTTT TCCGTGTTAA AAGATCATAT TAAGCAACAT
 601 GGTCAGCAAA ATGAAGTGAT ACTGATGTGC TCAGAGTGCC ATATTACATC
 651 TAGAAGCCAG GAGGAACTTG AAGCCCACGT GGTGAATGAC CATGACAATG
 701 ATGCCAATAT CCACACCCAA TCCAAAGCCC AACAGTGCGT AAGCCCCTCC
 751 AGCTCTTTGT GTCGGAAAAC CACAGAAAGA AATGAAACCA TTCCAGATAT
 801 CCCAGTAAGT GTGGACAATC TACAGACTCA TACTGTCCAA ACTGCATCTG
 851 TGGCAGAAAT GGGTAGGAGG AAATGGTATG CATACGAACA GTACGGCATG
 901 TATCGATGCT TGTTTTGTAG TTATACTTGT GGCCAGCAGA GAATGTTGAA
 951 AACACACGCT TGGAAACATG CTGGGGAGGT TGATTGCTCC TATCCAATCT
1001 TTGAAAATGA AAATGAACCC CTAGGCCTGC TGGATTCTTC AGCAGCTGCT
1051 GCGCCTGGTG GGGTCGATGC AGTCGTCATT GCTATTGGAG AGAGTGAACT
1101 GAGTATCCAC AATGGGCCAT CAGTGCAAGT GCAGATTTGC AGCTCAGAAC
1151 AGTTATCATC TTCATCTCCT TTAGAACAGA GTGCAGAAAG AGGAGTACAC
1201 CTAAGTCAGT CAGTTACCCT GGACCCCAAT GAGGAAGAAA TGCTAGAAGT
1251 GATTTCTGAT GCAGAGGAGA ATCTGATTCC TGATAGCCTG CTTACATCAG
1301 CACAGAAAAT CATCAGCAGC AGCCCCAATA AAAAAGGGCA TGTTAACGTG
1351 ATAGTGGAGC GATTGCCAAG TGCTGAAGAA ACCCTTTCAC AGAAGCGCTT
1401 CCTCATGAAC ACTGAAATGG AAGAAGGGAA GGACCTGAGC CTGACAGAAG
1451 CTCAGATTGG GCGCGAAGGA ATGGATGATG TTTATCGTGC TGATAAATGT
1501 ACTGTTGATA TTGGGGGATT GATCATAGGC TGGAGCAGTT CAGAGAAAAA
1551 AGACGAGTTA ATGAATAAAG GCCTGGCTAC TGATGAGAAT GCCCCACCAG
1601 GCCGGAGAAG GACAAATTCT GAGTCTCTTC GATTACACTC ATTAGCTGCA
1651 GAAGCCCTTG TCACAATGCC TATAAGAGCT GCAGAGTTGA CAAGAGCCAA
1701 CCTGGGGCAC TATGGAGATA TAAACCTTTT AGATCCAGAT ACTAGTCAAA
1751 GGCAAGTAGA TAGTACATTG GCAGCGTACT CAAAAATGAT GTCGCCACTT
1801 AAAAACTCTT CAGATGGATT AACTAGTCTT AACCAAAGCA ACTCCACCTT
1851 GGTAGCACTC CCAGAGGGTA GGCAGGAATT GTCAGATGGG CAGGTTAAGA
1901 CAGGCATCAG CATGTCCTTA CTCACCGTCA TTGAAAAATT GAGAGAAAGG
1951 ACAGACCAAA ACGCTTCAGA CGATGACATT TTGAAAGAGT TGCAGGACAA
2001 CGCCCAGTGC CAACCCAACA GCGATACAAG TTTGTCCGGA AACAATGTGG
2051 TGGAATACAT CCCGAATGCT GAACGACCCT ACCGTTGCCG CCTGTGTCAC
2101 TACACAAGTG GCAACAAGGG CTACATCAAG CAGCACTTAC GAGTCCATCG
2151 ACAGAGACAG CCTTATCAGT GTCCTATCTG CGAGCACATA GCGGACAACA
2201 GCAAAGATTT GGAGAGTCAC ATGATCCACC ACTGTAAGAC AAGAATATAC
2251 CAGTGCAAGC AGTGTGAAGA ATCCTTCCAT TATAAGAGTC AATTGAGGAA
2301 CCATGAGAGA GAACAGCACA GTCTTCCAGA TACCTTGTCA ATAGCAACTT
2351 CTAATGAGCC AAGAATTTCC AGTGATACAG CTGATGGAAA ATGTGTCCAG
2401 GAAGGGAATA AGTCTTCAGT CCAGAAACAA TATAGATGTG ATGTGTGTGA
2451 TTATACAAGT ACAACATATG TTGGTGTCAG AAACCACAGG CGAATCCATA
2501 ACTCTGATAA GCCGTACAGA TGCTCTCTGT GTGGGTATGT GTGTAGCCAT
2551 CCTCCTTCTT TGAAGTCTCA TATGTGGAAA CATGCAAGTG ACCAAAATTA
```

```
2601 CAACTACGAA CAAGTAAACA AGGCTATTAA CGACGCGATT TCACAAAGTG
2651 GCAGAGTTCT GGGGAAATCC CCTGGAAAGA CTCAATTAAA GAGCAGTGAA
2701 GAGAGTGCAG ATCCCGTCAC TGGAAGTTCG GAAAATGCAG TGTCATCTTC
2751 AGAACTGATG TCCCAGACTC CCAGTGAAGT TCTGGGTACC AACGAGAATG
2801 AGAAACTGAG CCCTACAAGT AATACCTCAT ATAGTTTAGA AAAAATCTCC
2851 AGTCTGGCCC CTCCTAGCAT GGAGTACTGC GTTTTACTCT TCTGCTGTTG
2901 TATTTGTGGT TTTGAATCAA CCAGCAAAGA AAACCTCTTG GATCATATGA
2951 AAGAGCACGA GGGTGAAATT GTAAACATCA TCCTGAATAA GGACCACAAT
3001 ACAGCTCTAA ACACAAATTA GGTGGAATAA TGACTCGAGC AGGAAAGCAG
3051 TAGAAGAGGA TTCCTTCACC ACAGTTTCAC CTTTACGCTG TCAGACAACT
3101 TCCTGCCACA GAAGAAGTCG TTGATGTGAT TTTTGAGGAA ATGACAGATG
3151 TGACTTTGGA ACCAAACTTG TAATAAAAGG AATTCCAAAT GGAAAAAAAA
3201 AAAAA
```

BLAST Results

No BLAST result

Medline entries

90301500:
Cloning and sequencing of a zinc finger cDNA expressed in mouse testis.

92310982:
Zfp-37, a new murine zinc finger encoding gene, is expressed in a developmentally regulated
pattern in the male germ line.

Peptide information for frame 1

ORF from 472 bp to 3018 bp; peptide length: 849
Category: similarity to known protein

```
  1 MSQTNFTPDT LAQNEGKAMS YQCSLCKFLS SSFSVLKDHI KQHGQQNEVI
 51 LMCSECHITS RSQEELEAHV VNDHDNDANI HTQSKAQQCV SPSSSLCRKT
101 TERNETIPDI PVSVDNLQTH TVQTASVAEM GRRKWYAYEQ YGMYRCLFCS
151 YTCGQQRMLK THAWKHAGEV DCSYPIFENE NEPLGLLDSS AAAAPGGVDA
201 VVIAIGESEL SIHNGPSVQV QICSSEQLSS SSPLEQSAER GVHLSQSVTL
251 DPNEEEMLEV ISDAEENLIP DSLLTSAQKI ISSSPNKKGH VNVIVERLPS
301 AEETLSQKRF LMNTEMEEGK DLSLTEAQIG REGMDDVYRA DKCTVDIGGL
351 IIGWSSSEKK DELMNKGLAT DENAPPGRRR TNSESLRLHS LAAEALVTMP
401 IRAAELTRAN LGHYGDINLL DPDTSQRQVD STLAAYSKMM SPLKNSSDGL
451 TSLNQSNSTL VALPEGRQEL SDGQVKTGIS MSLLTVIEKL RERTDQNASD
501 DDILKELQDN AQCQPNSDTS LSGNNVVEYI PNAERPYRCR LCHYTSGNKG
551 YIKQHLRVHR QRQPYQCPIC EHIADNSKDL ESHMIHHCKT RIYQCKQCEE
601 SFHYKSQLRN HEREQHSLPD TLSIATSNEP RISSDTADGK CVQEGNKSSV
651 QKQYRCDVCD YTSTTYVGVR NHRRIHNSDK PYRCSLCGYV CSHPPSLKSH
701 MWKHASDQNY NYEQVNKAIN DAISQSGRVL GKSPGKTQLK SSEESADPVT
751 GSSENAVSSS ELMSQTPSEV LGTNENEKLS PTSNTSYSLE KISSLAPPSM
801 EYCVLLFCCC ICGFESTSKE NLLDHMKEHE GEIVNIILNK DHNTALNTN
```

BLASTP hits

Entry S10245 from database PIR:
finger protein, testis - mouse
Score = 265, P = 8.4e-23, identities = 61/205, positives = 91/205

Entry S22954 from database PIR:
finger protein zfp-37 - mouse
Score = 265, P = 9.1e-22, identities = 61/205, positives = 91/205

Entry AF031657_1 from database TREMBL:
gene: "Zfp94"; product: "zinc-finger protein 94"; Rattus norvegicus zinc-finger protein 94 (Zfp94) gene, partial cds.
Score = 243, P = 1.6e-21, identities = 57/190, positives = 85/190

Alert BLASTP hits for DKFZphtes3_2e12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2e12, frame 1

Report for DKFZphtes3_2e12.1

```
[LENGTH]      849
[MW]          94325.42
[pI]          5.47
[HOMOL]       PIR:A54661 zinc finger protein ZNF41 - human (fragment) 2e-22
[FUNCAT]      04.05.01.04 transcriptional control  [S. cerevisiae, YJL056c] 3e-09
[FUNCAT]      30.10 nuclear organization    [S. cerevisiae, YJL056c] 3e-09
[FUNCAT]      04.03.01 trna synthesis       [S. cerevisiae, YPR186c PZF1 - TFIIIA] 1e-07
[FUNCAT]      04.01.01 rrna synthesis       [S. cerevisiae, YPR186c PZF1 - TFIIIA] 1e-07
[FUNCAT]      04.99 other transcription activities [S. cerevisiae, YOR113w] 4e-07
[FUNCAT]      01.05.04 regulation of carbohydrate utilization     [S. cerevisiae, YGL209w]
2e-04
[FUNCAT]      13.04 homeostasis of other ions      [S. cerevisiae, YNL027w] 2e-04
[FUNCAT]      11.01 stress response [S. cerevisiae, YMR037c] 3e-04
[BLOCKS]      BL00028 Zinc finger, C2H2 type, domain proteins
[SCOP]        d1meyg_ 9.6.1.1.1 a designed zinc finger protein [syntheti 8e-06
[PIRKW]       nucleus 8e-18
[PIRKW]       RNA binding 5e-13
[PIRKW]       duplication 7e-13
[PIRKW]       tandem repeat 1e-21
[PIRKW]       spermatogenesis 6e-16
[PIRKW]       zinc 9e-21
[PIRKW]       zinc finger 1e-21
[PIRKW]       DNA binding 1e-21
[PIRKW]       metal binding 3e-15
[PIRKW]       phosphoprotein 5e-13
[PIRKW]       leucine zipper 1e-13
[PIRKW]       alternative splicing 6e-18
[PIRKW]       eye lens 2e-16
[PIRKW]       oocyte 1e-12
[PIRKW]       transcription factor 6e-18
[PIRKW]       segmentation 7e-13
[PIRKW]       embryo 1e-12
[PIRKW]       transcription regulation 2e-19
[PIRKW]       homeobox 2e-08
[SUPFAM]      POZ domain homology 7e-15
[SUPFAM]      transcription factor Krueppel 7e-13
[SUPFAM]      zinc finger protein ZFP-36 1e-21
[SUPFAM]      homeobox homology 2e-08
[SUPFAM]      unassigned homeobox proteins 2e-08
[PROSITE]     CYTOCHROME_C    1
[PROSITE]     MYRISTYL        10
[PROSITE]     ZINC_FINGER_C2H2       3
[PROSITE]     AMIDATION       2
[PROSITE]     CAMP_PHOSPHO_SITE      2
[PROSITE]     CK2_PHOSPHO_SITE       18
[PROSITE]     TYR_PHOSPHO_SITE       3
[PROSITE]     PKC_PHOSPHO_SITE       10
[PROSITE]     ASN_GLYCOSYLATION      7
[PFAM]        Zinc finger, C2H2 type
[KW]          Irregular
[KW]          3D
[KW]          LOW_COMPLEXITY     5.65 %

SEQ    MSQTNFTPDTLAQNEGKAMSYQCSLCKFLSSSFSVLKDHIKQHGQQNEVILMCSECHITS
SEG    ......................xxxxxxxxxxxxxxx.......................
1meyF  ............................................................

SEQ    RSQEELEAHVVNDHDNDANIHTQSKAQQCVSPSSSLCRKTTERNETIPDIPVSVDNLQTH
SEG    ............................................................
1meyF  ............................................................

SEQ    TVQTASVAEMGRRKWYAYEQYGMYRCLFCSYTCGQQRMLKTHAWKHAGEVDCSYPIFENE
SEG    ............................................................
1meyF  ............................................................

SEQ    NEPLGLLDSSAAAAPGGVDAVVIAIGESELSIHNGPSVQVQICSSEQLSSSSPLEQSAER
SEG    .......xxxxxxxxxxxxxxxxxxx.................xxxxxxxxxxxxxx...
1meyF  ............................................................

SEQ    GVHLSQSVTLDPNEEEMLEVISDAEENLIPDSLLTSAQKIISSSPNKKGHVNVIVERLPS
SEG    ............................................................
1meyF  ............................................................
```

```
SEQ     AEETLSQKRFLMNTEMEEGKDLSLTEAQIGREGMDDVYRADKCTVDIGGLIIGWSSSEKK
SEG     ............................................................
1meyF   ............................................................

SEQ     DELMNKGLATDENAPPGRRRTNSESLRLHSLAAEALVTMPIRAAELTRANLGHYGDINLL
SEG     ............................................................
1meyF   ............................................................

SEQ     DPDTSQRQVDSTLAAYSKMMSPLKNSSDGLTSLNQSNSTLVALPEGRQELSDGQVKTGIS
SEG     ............................................................
1meyF   ............................................................

SEQ     MSLLTVIEKLRERTDQNASDDDILKELQDNAQCQPNSDTSLSGNNVVEYIPNAERPYRCR
SEG     ............................................................
1meyF   .....................................................TTTEETT

SEQ     LCHYTSGNKGYIKQHLRVHRQRQPYQCPICEHIADNSKDLESHMIHHCKTRIYQCKQCEE
SEG     ............................................................
1meyF   TTTCEETTHHHHHHHHHHHHTTCCEEETTTTEEECCHHHHHHHHHHHHCCCCEEETTTTE

SEQ     SFHYKSQLRNHEREQHSLPDTLSIATSNEPRISSDTADGKCVQEGNKSSVQKQYRCDVCD
SEG     ............................................................
1meyF   EECCHHHHHHHHHHHC............................................

SEQ     YTSTTYVGVRNHRRIHNSDKPYRCSLCGYVCSHPPSLKSHMWKHASDQNYNYEQVNKAIN
SEG     ............................................................
1meyF   ............................................................

SEQ     DAISQSGRVLGKSPGKTQLKSSEESADPVTGSSENAVSSSELMSQTPSEVLGTNENEKLS
SEG     ............................................................
1meyF   ............................................................

SEQ     PTSNTSYSLEKISSLAPPSMEYCVLLFCCCICGFESTSKENLLDHMKEHEGEIVNIILNK
SEG     ............................................................
1meyF   ............................................................

SEQ     DHNTALNTN
SEG     .........
1meyF   .........
```

Prosite for DKFZphtes3_2e12.1

PS00001	104->108	ASN_GLYCOSYLATION	PDOC00001
PS00001	445->449	ASN_GLYCOSYLATION	PDOC00001
PS00001	454->458	ASN_GLYCOSYLATION	PDOC00001
PS00001	457->461	ASN_GLYCOSYLATION	PDOC00001
PS00001	497->501	ASN_GLYCOSYLATION	PDOC00001
PS00001	646->650	ASN_GLYCOSYLATION	PDOC00001
PS00001	784->788	ASN_GLYCOSYLATION	PDOC00001
PS00004	98->102	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	59->62	PKC_PHOSPHO_SITE	PDOC00005
PS00005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS00005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS00005	357->360	PKC_PHOSPHO_SITE	PDOC00005
PS00005	385->388	PKC_PHOSPHO_SITE	PDOC00005
PS00005	425->428	PKC_PHOSPHO_SITE	PDOC00005
PS00005	678->681	PKC_PHOSPHO_SITE	PDOC00005
PS00005	696->699	PKC_PHOSPHO_SITE	PDOC00005
PS00005	726->729	PKC_PHOSPHO_SITE	PDOC00005
PS00005	817->820	PKC_PHOSPHO_SITE	PDOC00005
PS00006	62->66	CK2_PHOSPHO_SITE	PDOC00006
PS00006	106->110	CK2_PHOSPHO_SITE	PDOC00006
PS00006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS00006	232->236	CK2_PHOSPHO_SITE	PDOC00006
PS00006	262->266	CK2_PHOSPHO_SITE	PDOC00006
PS00006	300->304	CK2_PHOSPHO_SITE	PDOC00006
PS00006	314->318	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS00006	381->385	CK2_PHOSPHO_SITE	PDOC00006
PS00006	485->489	CK2_PHOSPHO_SITE	PDOC00006
PS00006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS00006	617->621	CK2_PHOSPHO_SITE	PDOC00006
PS00006	626->630	CK2_PHOSPHO_SITE	PDOC00006
PS00006	741->745	CK2_PHOSPHO_SITE	PDOC00006
PS00006	758->762	CK2_PHOSPHO_SITE	PDOC00006
PS00006	766->770	CK2_PHOSPHO_SITE	PDOC00006
PS00006	817->821	CK2_PHOSPHO_SITE	PDOC00006

PS00007	331->339	TYR_PHOSPHO_SITE	PDOC00007
PS00007	703->711	TYR_PHOSPHO_SITE	PDOC00007
PS00007	596->605	TYR_PHOSPHO_SITE	PDOC00007
PS00008	142->148	MYRISTYL	PDOC00008
PS00008	185->191	MYRISTYL	PDOC00008
PS00008	196->202	MYRISTYL	PDOC00008
PS00008	241->247	MYRISTYL	PDOC00008
PS00008	349->355	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	478->484	MYRISTYL	PDOC00008
PS00008	645->651	MYRISTYL	PDOC00008
PS00008	751->757	MYRISTYL	PDOC00008
PS00008	772->778	MYRISTYL	PDOC00008
PS00009	130->134	AMIDATION	PDOC00009
PS00009	376->380	AMIDATION	PDOC00009
PS00028	146->167	ZINC_FINGER_C2H2	PDOC00028
PS00028	684->705	ZINC_FINGER_C2H2	PDOC00028
PS00028	595->617	ZINC_FINGER_C2H2	PDOC00028
PS00190	53->59	CYTOCHROME_C	PDOC00169

Pfam for DKFZphtes3_2e12.1

```
HMM_NAME       Zinc finger, C2H2 type

HMM                  *CpwPDCgKtFrrwsNLrRHMR.T.H*
                      C++  C+ T R++++L++H      H
Query           53   CSE--CHITSRSQEELEAHVVN-DH      74

23.25  (bits) f:  539 t:  559 Target: dkfzphtes3_2e12.1 similarity to finger proteins
  Alignment to HMM consensus:
Query                *CpwPDCgKtFrrwsNLrRHMRTH*
                      C    C++T  ++    ++H+R+H
  dkfzphtes3   539   CRL--CHYTSGNKGYIKQHLRVH     559

Query          f:  567 t:  587 Target: dkfzphtes3_2e12.1 similarity to finger proteins
  Alignment to HMM consensus:
HMM                  *CpwPDCgKtFrrwsNLrRHMRTH*
                      CP+  C+     ++ +L+ HM+ H
Query          567   CPI--CEHIADNSKDLESHMIHH     587

33.47  (bits) f:  595 t:  616 Target: dkfzphtes3_2e12.1 similarity to finger proteins
  Alignment to HMM consensus:
Query                *CpwPDCgKtFrrwsNLrRHMR.T.H*
                      C+   C+++F ++S+LR+H R    H
  dkfzphtes3   595   CKQ--CEESFHYKSQLRNHERE-QH     616

Query          f:  656 t:  676 Target: dkfzphtes3_2e12.1 similarity to finger proteins
  Alignment to HMM consensus:
HMM                  *CpwPDCgKtFrrwsNLrRHMRTH*
                      C++  C++T ++     R+H+R+H
Query          656   CDV--CDYTSTTYVGVRNHRRIH     676

24.53  (bits) f:  684 t:  704 Target: dkfzphtes3_2e12.1 similarity to finger proteins
  Alignment to HMM consensus:
Query                *CpwPDCgKtFrrwsNLrRHMRTH*
                      C+   CG++ +++ +L+ HM  H
  dkfzphtes3   684   CSL--CGYVCSHPPSLKSHMWKH     704

Query          f:  809 t:  829 Target: dkfzphtes3_2e12.1 similarity to finger proteins
  Alignment to HMM consensus:
HMM                  *CpwPDCgKtFrrwsNLrRHMRTH*
                      C +  CG    ++++NL   HM+ H
Query          809   CCI--CGFESTSKENLLDHMKEH     829
```

DKFZphtes3_2f14
----

group: testes derived

DKFZphtes3_2f14 encodes a novel 129 amino acid protein with very weak similarity to human omega protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to omega protein

complete cDNA, complete cds, 1 EST hit

Sequenced by EMBL

Locus: unknown

Insert length: 2353 bp
Poly A stretch at pos. 2341, no polyadenylation signal found

```
   1 GCAGATTCTC CAGGCCCAGC ATCTGCCTCA CCGTGGCCCC CCACAAGCCA
  51 AGCGCCTGCC TTTCAGCAGC CTCTACACAC CCAGCTCCTG CCACCCAATG
 101 GCTCTTTAGG CCAAGCTCAT ACCTCACGAT GATTTTTCCA GGCCCAACTT
 151 TTGTCTCATG GCAACCTTCC CTGGCCAAGT TTCCACCTAT TTCCTGGCAG
 201 CCTGGACAGG CCCAGGTCCT GCCACACACT GGCCTCTCTA CGCCCAGCTC
 251 ATGCCTCACA GTGGCCTCTC CAGGCCCAGC TCCTGTCCCG GGACATCATC
 301 TCCAGGCCCA AAACTTCCTC AAGTCGGCCT CTCCAGGCCC AGTTGCTGCC
 351 TCCCGGCATT CTCTCCAGGC CTAGCTCTTC CTCCTGGCTG TATCTACAAG
 401 ACCAACTCCT GCCTCACAAC AACCTTTTAT GGCTCAGCTC CTGCCCAACT
 451 ACTGCCGGCC TTTGTAGGCC CAAAACTTCC TCAAGTCAAG CTCTTTAGGC
 501 CCACCTTCTG CCTTGCAGTG GCCTGTACAG ACCCAGCTCT GGCTTGAGAA
 551 CAGCCTCTGC AGGCCCTGCT CTTGCCTCTT AGCTCCCTCT CCAGGCCCAT
 601 CTCTTGCCTC ACAGTGGCTT CCGTGGGCCA AGTTCCCGCC TGCCTCCCAG
 651 CAGCCTCAAC AGGCCTAGCT CCTCCCTCAC AATGGCTTGT TTAGGTCCAG
 701 TTGATGCCTC TGGCAACCTG TCCAGGCCCA GCTCCTGCCT CACACTGGCC
 751 TCTCTAGGCC GAGGTCCTTT CTCATACTGG CCTGTTTAGG CCCAGCTCAT
 801 TCCTCTTGTC ATCTCTCCAG GCCCAGCTTT TGCCTGTTGT TGGCCTCTAC
 851 CTCACAGTGC ACCTTCCAGT CCCACCTCTT GCCTCACCAT GGCCTCCTCT
 901 GACCAGGTTC CTGCCTTTCG GCAGCCTCTA CAGGCCTAGC TGCTGCCTCC
 951 CAATGGCCTT TGTAGGCCAC GCTCATGCCT CACTGTGGCC TTTCCAGGCC
1001 TAGCTTTCGC TTTTTGGCCA CTCCAGGCCC AGAACTTCCC CCAGTCAGCC
1051 TCTCCAGGCC CAGCTCTTCC TCCCAGCAAC CTCTGCAGGC CCAAATCATC
1101 CTCAAATTGG CCTCTTCTTT CCCAGCTCCT GCCTCCTGGT GGCCTCTGAA
1151 GACCCAAATC GTCCTCCAGT TGGTTTTTCC AGGCCCAGCT CCTGCCTTTT
1201 GGTGGCCTCT CCAGGTGCAA AACTTCCTCC CATCAGCCTG TCCAGGCCCA
1251 GCTCATGCCT CTTGGTGGCC TTCTCAGGCC CTGCTTTTGA CTTGGTGGCC
1301 TCTTCAGGCC CAGAACTTGA ACTCAAGTCA GCCTCTCCAG GCCCAGCTCC
1351 TGCCTTCTTA AGGTCTGTAC AGGCCCAGCC TCTACCTCAC AGCGGACTCT
1401 CCACACCCAG CTCTTGCCTC ACTGTAGCCT CCCCAGTCCA AAACTCCTGC
1451 CTTTTGGCAG CTTCGACAAG CCCAGCTCCT GCCTTTCAAT GACCTCTTTA
1501 GGCCCCGCTC ATTCCTTACA ACGGCCTTTC CAGGCCCAGT TTTTCCCTTT
1551 TGGCGGCCTC TCCAGGCCCA GAACTTCCTC AAGTCGGCCT CTTTAGGCCC
1601 AGTTGCTGCC TCCTGGCATC CTCTGCAGGC CGAGCTCTTC CTCCCTGCTG
1651 TGTCTACAGG CCCAACTCCT GCCTCACAAC AACCTCCTTG GACTCAGCTT
1701 CTGCCCAGCT CCTGGTGGCC TTTGTAGGCT CAAAATTTTC TCAAATCAAG
1751 CTCTCCAGGC CTACTGTCAG CCTCGTGGCA GCCTAAACAG GCCCAGCTCC
1801 TGCCTGACAA TGGCCTCTCC AGGCTTTTCT CCTGCCTCGC AGCAGGCTTT
1851 CCAGGCCCAG CTCTTGCCTC ATGGTGGCCT TCCCCGGCCA TGTTCCTATC
1901 TGACTTCTGG CAGCCTCAAC CGGCCCAGCT TCTGCCTCAC ACTGGCCTCT
1951 CTAGGCCCAG CTCCTTTTTC ACAGTGGCCT CACTACGCCC ATCTCCTACC
2001 TCAGATCTGC CTCCCAAGAC CCAGCTCCTG TCTCATGGTG GTCTCTCTTA
2051 CACCAGCTCC TGCCTCACAA TGGCCTCGTC TGGCCCATCT TCTGCCTCAC
2101 AGTGGCCACT CAAGGCCCAT CTTTTGCCTC ATGGTAGCCT CTTCTGGTTT
2151 TGCTCTTGCC TCACAGTTGC CTCTTCCAGA TCCAGCTTTA AGCCTTTGAT
2201 GGTCAACAGC ATCAAGGAGC CTAAAGCTTC CCTGGACTCT CATTTGTTCA
2251 CTTTACAGCA GAGTGCCTTA GCAAAAACTG TCTCTTAACC TTGAGAGTGG
2301 ATTTCTGACA AATCGATAGT AAATTCTGCC TGTGTGGTTT CAAAAAAAAA
2351 AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 158 bp to 544 bp; peptide length: 129
Category: similarity to known protein

```
  1 MATFPGQVST YFLAAWTGPG PATHWPLYAQ LMPHSGLSRP SSCPGTSSPG
 51 PKLPQVGLSR PSCCLPAFSP GLALPPGCIY KTNSCLTTTF YGSAPAQLLP
101 AFVGPKLPQV KLFRPTFCLA VACTDPALA
```

BLASTP hits

Entry I70697 from database PIR:
omega protein - human (fragment)
Score = 79, P = 2.8e-03, identities = 32/94, positives = 38/94

Alert BLASTP hits for DKFZphtes3_2f14, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2f14, frame 2

Report for DKFZphtes3_2f14.2

```
[LENGTH]     129
[MW]         13421.76
[pI]         9.14
[PROSITE]    MYRISTYL      2
[KW]         Irregular
[KW]         LOW_COMPLEXITY    10.85 %
```

```
SEQ    MATFPGQVSTYFLAAWTGPGPATHWPLYAQLMPHSGLSRPSSCPGTSSPGPKLPQVGLSR
SEG    ...................................xxxxxxxxxxxxxxx.........
PRD    cccccccceeehhhhhcccccccccccccccccccccccccccccccccccccccccccc

SEQ    PSCCLPAFSPGLALPPGCIYKTNSCLTTTFYGSAPAQLLPAFVGPKLPQVKLFRPTFCLA
SEG    ............................................................
PRD    cccccccccccccccccccccccccceeeccccccccccccccccccccccccccccccc

SEQ    VACTDPALA
SEG    .........
PRD    ccccccccc
```

Prosite for DKFZphtes3_2f14.2

```
PS00008     6->12    MYRISTYL          PDOC00008
PS00008    92->98    MYRISTYL          PDOC00008
```

(No Pfam data available for DKFZphtes3_2f14.2)

DKFZphtes3_2g7
----

group: testes derived

DKFZphtes3_2g7 encodes a novel 359 amino acid protein with similarity to neurofiliament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

complete cDNA, complete cds, 6 EST hits (5 hits are out of a testis library)

Sequenced by EMBL

Locus: unknown

Insert length: 1613 bp
Poly A stretch at pos. 1595, polyadenylation signal at pos. 1557

```
   1 GCCACACAGG CTCCTTGGAG TAAGAGTGTG AGAAACTGGA TGAAGACAGC
  51 TGTATTCTTT TGGAAGCGTT CGAGATTGGT CTGTCTCTAC CAACTAAAAA
 101 CTTCTAGCTT AAGTGCAGAG ATTTAAGGAG ATCAACAAAA ACTCAGTCTA
 151 GACATATTAT GAGGCTGGGA GGGTATCAAC AGACTTGAGT TCTTGTCAGC
 201 AAGATCACCT GCTTTTAATA TTGTCCTCAG GGTCTGAGCA CATCTGGAAG
 251 TGAGGTCAAT CAAGTTAGAC CCCAAAAACT TTTGTGACAA CAGTGAAGAG
 301 GGGAAAATAA ACACACCACA AACATGAACC TCAACCCCCC GACATCTGCT
 351 CTTCAGATCG AGGGCAAAGG CAGCCATATT ATGGCTAGAA ATGTAAGCTG
 401 CTTTCTAGTC AGGCACACCC CTCATCCCAG AAGAGTCTGC CACATCAAAG
 451 GCTTGAATAA CATTCCAATC TGTACTGTGA ATGATGATGA GAATGCATTT
 501 GGAACATTGT GGGAAGTTGG CCAGTCTAAC TACTTAGAGA AGAACAGGAT
 551 ACCATTTGCC AATTGCAGTT ACCCCCCGAG CACTGCAGTC CAGAAGAGCC
 601 CTGTAAGAGG AATGTCGCCA GCCCCAAACG GTGCCAAAGT GCCTCCACGG
 651 CCTCATTCTG AGCCCAGTAG AAAAATTAAA GAGTGCTTCA AAACTTCCAG
 701 TGAGAATCCC TTAGTAATTA AAAAGGAAGA AATTAAGGCC AAAAGACCAC
 751 CATCACCTCC AAAGGCATGC TCTACTCCTG GCTCCTGTTC TTCAGGGATG
 801 ACAAGTACCA AGAATGATGT GAAAGCAAAC ACCATTTGCA TACCAAACTA
 851 TCTGGATCAG GAAATAAAAA TCCTGGCAAA GCTCTGTAGC ATTTTGCATA
 901 CTGATTCTCT GGCAGAAGTT TTACAGTGGC TGCTTCATGC AACTTCAAAA
 951 GAAAAAGAGT GGGTCTCAGC TTTGATTCAT TCTGAGCTTG CCGAGATAAA
1001 CCTGTTAACT CATCACAGAA GAAACACCTC AATGGAACCA GCAGCAGAGA
1051 CTGGGAAGCC ACCCACAGTT AAATCACCAC CCACAGTTAA ATTGCCCCCA
1101 AATTTTACTG CAAAATCAAA AGTGCTGACC AGAGATACAG AAGGGGATCA
1151 ACCAACCAGA GTGTCAAGTC AAGGATCTGA AGAAAACAAG GAAGTACCAA
1201 AAGAGGCTGA GCACAAGCCT CCACTACTTA TAAGAAGAAA TAATATGAAA
1251 ATACCTGTTG CAGAATATTT CAGCAAACCA AATTCTCCTC CCAGGCCTAA
1301 CACTCAGGAG AGTGGATCAG CAAAACCAGT GTCAGCAAGG AGTATACAAG
1351 AATACAACCT CTGTCCCCAA AGAGCATGTT ATCCTTCAAC ACACCGGAGG
1401 TAGAAGTTCT AGACTGGGTG AATTCTTTCA TGAATATGAG CTTCACATTT
1451 ACATCATCAA ATTATTTTTC AAATGAATAT TTTTGGTATT GAGGAATCAA
1501 GTGGTCCTCT TTATGGTGGC ACATGTAAAT CTAAAAATAC CTGTATGTAA
1551 TGCTACAAAT AAATATTACT GGAAATGATA TTTCCATTTG TAGTTAAAAA
1601 AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 324 bp to 1400 bp; peptide length: 359
Category: similarity to known protein

```
  1 MNLNPPTSAL QIEGKGSHIM ARNVSCFLVR HTPHPRRVCH IKGLNNIPIC
 51 TVNDDENAFG TLWEVGQSNY LEKNRIPFAN CSYPPSTAVQ KSPVRGMSPA
101 PNGAKVPPRP HSEPSRKIKE CFKTSSENPL VIKKEEIKAK RPPSPPKACS
151 TPGSCSSGMT STKNDVKANT ICIPNYLDQE IKILAKLCSI LHTDSLAEVL
201 QWLLHATSKE KEWVSALIHS ELAEINLLTH HRRNTSMEPA AETGKPPTVK
251 SPPTVKLPPN FTAKSKVLTR DTEGDQPTRV SSQGSEENKE VPKEAEHKPP
301 LLIRRNNMKI PVAEYFSKPN SPPRPNTQES GSAKPVSARS IQEYNLCPQR
351 ACYPSTHRR
```

BLASTP hits

Entry A43427 from database PIR:
neurofilament triplet H1 protein - rabbit (fragment)
Score = 118, P = 5.6e-04, identities = 79/290, positives = 110/290

Entry RNNFH_1 from database TREMBL:
Rat high molecular weight neurofilament (NF-H) protein mRNA, 3' end.
Score = 115, P = 9.5e-04, identities = 69/281, positives = 100/281

Entry B43427 from database PIR:
neurofilament protein H form H2 (repetitive region) - rabbit (fragment)
Score = 111, P = 1.3e-03, identities = 64/269, positives = 102/269

Alert BLASTP hits for DKFZphtes3_2g7, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2g7, frame 3

Report for DKFZphtes3_2g7.3

[LENGTH]	359	
[MW]	39725.53	
[pI]	9.45	
[PROSITE]	MYRISTYL	3
[PROSITE]	CAMP_PHOSPHO_SITE	1
[PROSITE]	CK2_PHOSPHO_SITE	9
[PROSITE]	PKC_PHOSPHO_SITE	10
[PROSITE]	ASN_GLYCOSYLATION	4
[KW]	Alpha_Beta	
[KW]	LOW_COMPLEXITY	4.18 %

```
SEQ  MNLNPPTSALQIEGKGSHIMARNVSCFLVRHTPHPRRVCHIKGLNNIPICTVNDDENAFG
SEG  ............................................................
PRD  cccccccceeeccccccceeeecccceeeeecccccccccccccccccccccccccccccc

SEQ  TLWEVGQSNYLEKNRIPFANCSYPPSTAVQKSPVRGMSPAPNGAKVPPRPHSEPSRKIKE
SEG  ............................................................
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhh

SEQ  CFKTSSENPLVIKKEEIKAKRPPSPPKACSTPGSCSSGMTSTKNDVKANTICIPNYLDQE
SEG  ............................................................
PRD  hccccccccceeeehhhhhhcccccccccccccccccccccccccccceeeecccccchhh

SEQ  IKILAKLCSILHTDSLAEVLQWLLHATSKEKEWVSALIHSELAEINLLTHHRRNTSMEPA
SEG  ............................................................
PRD  hhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccc

SEQ  AETGKPPTVKSPPTVKLPPNFTAKSKVLTRDTEGDQPTRVSSQGSEENKEVPKEAEHKPP
SEG  ....xxxxxxxxxxxxxx..........................................
PRD  ccccccccccccccccccccccceeeeeccccccccceeeecccccccccccccccccccc

SEQ  LLIRRNNMKIPVAEYFSKPNSPPRPNTQESGSAKPVSARSIQEYNLCPQRACYPSTHRR
SEG  ...........................................................
PRD  eeeecccccccceeeeeccccccccccccccccccccchhhhhhccccccccccccccccc
```

Prosite for DKFZphtes3_2g7.3

PS00001	23->27	ASN_GLYCOSYLATION	PDOC00001
PS00001	80->84	ASN_GLYCOSYLATION	PDOC00001
PS00001	234->238	ASN_GLYCOSYLATION	PDOC00001

PS00001	260->264	ASN_GLYCOSYLATION	PDOC00001
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	161->164	PKC_PHOSPHO_SITE	PDOC00005
PS00005	207->210	PKC_PHOSPHO_SITE	PDOC00005
PS00005	243->246	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	262->265	PKC_PHOSPHO_SITE	PDOC00005
PS00005	332->335	PKC_PHOSPHO_SITE	PDOC00005
PS00005	337->340	PKC_PHOSPHO_SITE	PDOC00005
PS00005	356->359	PKC_PHOSPHO_SITE	PDOC00005
PS00006	51->55	CK2_PHOSPHO_SITE	PDOC00006
PS00006	61->65	CK2_PHOSPHO_SITE	PDOC00006
PS00006	124->128	CK2_PHOSPHO_SITE	PDOC00006
PS00006	162->166	CK2_PHOSPHO_SITE	PDOC00006
PS00006	195->199	CK2_PHOSPHO_SITE	PDOC00006
PS00006	207->211	CK2_PHOSPHO_SITE	PDOC00006
PS00006	235->239	CK2_PHOSPHO_SITE	PDOC00006
PS00006	272->276	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	158->164	MYRISTYL	PDOC00008
PS00008	284->290	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_2g7.3)

DKFZphtes3_2h1

group: transmembrane protein

DKFZphtes3_2h1 encodes a novel 116 amino acid protein with weak similarity to C. elegans cosmid C13F10.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to C.elegans C13F10.5

TRANSMEMBRANE 1

Sequenced by EMBL

Locus: /map="2"

Insert length: 1156 bp
Poly A stretch at pos. 1143, polyadenylation signal at pos. 1121

```
   1 GGCCATCAAA ATAACTAAAC CATGTCATTT GGAGCAACAA AGCCACTGCG
  51 GCCTCCATTT GGGCCAAGCT CTGACTGCAA TGATGCCTCT GCCCCGACCC
 101 GGGCCTCGCT GTGACTGACA ATGCCGCTGC ATCTTTTCAG CAGTCATTGA
 151 TGAGGAAGTA TCTACATCCT CCTTCCCACT ACCAGATTTT GCTTGGAGAA
 201 AAGCAGTTTC CTGAAATAAT TCTGTGACGA GCTTCTTCCA CATTAGGACA
 251 AAAATGCTGG AAGCGGCTCA GCCCCAGGGC AGCACATCAG AGACACCATG
 301 GAACACAGCC ATTCCTCTGC CGTCGTGCTG GGACCAGTCT TTCCTGACCA
 351 ATATCACCTT CTTGAAGGTT CTTCTCTGGT TGGTCCTGCT GGGACTGTTT
 401 GTGGAACTGG AATTTGGCCT GGCATATTTT GTCCTGTCCT TGTTCTATTG
 451 GATGTACGTC GGGACACGAG GCCCTGAAGA GAAGAAAGAG GGAGAGAAGA
 501 GCGCCTACTC TGTGTTCAAT CCAGGCTGTG AAGCCATCCA GGGCACCCTG
 551 ACTGCAGAGC AGTTGGAGCG CGAGTTACAG TTGAGACCCC TGGCAGGGAG
 601 ATAGGACCCA GCTGTGCTGT CATGCAGCTA ACCTCTGATG TGGTCTTCCT
 651 CACCATTGGC TATGGATTTG ATTTCAGGTG TATAGGACTA AGGGCAGCTT
 701 GCGGGTTAGC TCTGTGACTG CATAGTTTTT CTACCTTCTT TCCCTGATCT
 751 TTTGCTGCCA TTTGATCTTT GATAGTTTTG GTGAAACTCT CTAAAATACA
 801 TTCACTGTGG GTCCGACGCA ATTTATAAAA ATTATGTACT CAAGAAGGGA
 851 GACCTGTTTG TTTCATTTCT CATCTGTTTG GGAGATGATT TTAGAGCACT
 901 AGAAAGGCAC TGGGGAGATT CTCAGCTTAA AACATCCAGC AGTTTGAAGT
 951 ATGATTAGGT ACATCAGGGC TGCATTGTCA ATGTTCTCTT TAAGTCTTTT
1001 AACATTTATA GCAATTTTTT TTTTCCCGGA GAGTTTAGGT TGCAAGTTTT
1051 GGGTTTCTTG TTTGTTTTTG TTTTGCTTCC TGCTTTAATT CTTTAATTTT
1101 CAGTCATTAC TGGTATTGAA AAATAAAATA TCTTTAAAAC ATCAAAAAAA
1151 AAAAAA
```

BLAST Results

Entry HS313307 from database EMBL:
human STS SHGC-16715.
Score = 1222, P = 1.4e-48, identities = 248/251

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 254 bp to 601 bp; peptide length: 116
Category: similarity to unknown protein

```
  1 MLEAAQPQGS TSETPWNTAI PLPSCWDQSF LTNITFLKVL LWLVLLGLFV
 51 ELEFGLAYFV LSLFYWMYVG TRGPEEKKEG EKSAYSVFNP GCEAIQGTLT
101 AEQLERELQL RPLAGR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2h1, frame 2

TREMBL:CEUC13F10_2 gene: "C13F10.5"; Caenorhabditis elegans cosmid C13F10., N = 1, Score = 141, P = 8.2e-10

```
>TREMBL:CEUC13F10_2 gene: "C13F10.5"; Caenorhabditis elegans cosmid
     C13F10.
          Length = 171

  HSPs:

 Score = 141 (21.2 bits), Expect = 8.2e-10, P = 8.2e-10
 Identities = 32/82 (39%), Positives = 52/82 (63%)

Query:    27 DQSFLTNITFLKVLLWLVLLGLFVELEFGLAYFVLSLFYWMYVGTRGPEEKKEGEKSAYS 86
             +QS ++   T + V++++V    L    ++FG  +F+LSL  + Y  T G    ++ GE SAYS
Sbjct:    90 EQSVVS--TRIAVVVYVVGQALAAWVQFGAVFFILSLILFTYWNT-G--RRRRGEMSAYS 144

Query:    87 VFNPGCEAIQGTLTAEQLEREL 108
             VFN  CE + G++TAE  ER++
Sbjct:   145 VFNDNCERLAGSMTAEHFERDM 166
```

Pedant information for DKFZphtes3_2h1, frame 2

Report for DKFZphtes3_2h1.2

[LENGTH]	116	
[MW]	13092.19	
[pI]	4.64	
[PROSITE]	MYRISTYL	1
[PROSITE]	CK2_PHOSPHO_SITE	2
[PROSITE]	TYR_PHOSPHO_SITE	2
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	TRANSMEMBRANE	1
[KW]	LOW_COMPLEXITY	32.76 %

```
SEQ   MLEAAQPQGSTSETPWNTAIPLPSCWDQSFLTNITFLKVLLWLVLLGLFVELEFGLAYFV
SEG   ...................................xxxxxxxxxxxxxxxxxxxxx....
PRD   cccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhchhhhh
MEM   .................................MMMMMMMMMMMMMMMMMM..........

SEQ   LSLFYWMYVGTRGPEEKKEGEKSAYSVFNPGCEAIQGTLTAEQLERELQLRPLAGR
SEG   ......................................xxxxxxxxxxxxxxxxxx..
PRD   hhhhhhhhccccchhhhhcccceeeeccccccccccccchhhhhhhhhhccccccc
MEM   ........................................................
```

Prosite for DKFZphtes3_2h1.2

PS00001	33->37	ASN_GLYCOSYLATION	PDOC00001
PS00006	10->14	CK2_PHOSPHO_SITE	PDOC00006
PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00007	78->86	TYR_PHOSPHO_SITE	PDOC00007
PS00007	77->86	TYR_PHOSPHO_SITE	PDOC00007
PS00008	97->103	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_2h1.2)

DKFZphtes3_2h15
----

group: testes derived

DKFZphtes3_2h15 encodes a novel 855 amino acid protein with very weak similarity to S. pombe cdc23.

No informative BLAST results; no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to cdc23

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 4619 bp
Poly A stretch at pos. 4598, polyadenylation signal at pos. 4589

```
   1 GAAGGCGTCC CGGCATCGGC CAAGATTCTA CATTGCTCAT CTGGGCATCT
  51 GAGCCTCCTT CGAAGTTTCC TGTCACAACT GTCCTCTTGA CAGCATGGAT
 101 GAGGAGGAAG ACAATCTGTC TCTGCTGACC GCACTGCTGG AAGAAAATGA
 151 GTCAGCCTTG GATTGTAATT CAGAAGAAAA TAACTTCTTG ACGCGGGAAA
 201 ATGGCGAGCC CGACGCATTT GATGAGCTCT TTGATGCCGA CGGCGACGGT
 251 GAATCTTATA CAGAAGAGGC TGATGATGGA GAAACAGGAG AGACAAGAGA
 301 CGAAAAGGAA AATCTGGCCA CTCTCTTTGG AGATATGGAG GACTTAACAG
 351 ATGAAGAAGA AGTTCCCGCA TCACAGTCAA CTGAAAATAG GGTCCTCCCT
 401 GCTCCTGCCC CCAGGCGAGA GAAAACGAAT GAAGAGTTGC AAGAGGAATT
 451 AAGGAATTTG CAAGAGCAAA TGAAGGCCTT ACAAGAGCAG CTAAAAGTAA
 501 CAACAATTAA ACAGACAGCA AGCCCAGCCC GTCTGCAAAA ATCCCCTGAG
 551 AAGTCTCCCC GGCCACCTCT TAAGGAGAGG AGAGTTCAGA GAATTCAGGA
 601 GTCAACATGC TTTTCTGCGG AGCTTGATGT CCCTGCGCTA CCAAGAACCA
 651 AGAGGGTGGC TCGAACACCA AAGCCTTCAC CTCCAGATCC CAAAAGCTCA
 701 TCTTCAAGGA TGACAAGTGC ACCCTCCCAA CCCCTACAGA CGATTTCTCG
 751 GAACAAACCT AGTGGGATAA CTAGAGGTCA AATTGTGGGG ACCCCAGGAA
 801 GTTCTGGGGA AACGACTCAA CCCATCTGTG TGGAAGCCTT CTCTGGTCTG
 851 CGGCTCAGGC GGCCTCGAGT ATCCTCCACA GAAATGAACA AGAAAATGAC
 901 CGGCCGAAAA CTGATCAGAC TGTCTCAGAT CAAGGAAAAG ATGGCCAGAG
 951 AGAAGCTGGA AGAAATAGAT TGGGTGACAT TTGGGGTTAT ATTGAAGAAG
1001 GTTACGCCAC AGAGTGTGAA TAGTGGAAAA ACCTTCAGCA TATGGAAACT
1051 GAATGATCTT CGTGACCTGA CACAATGTGT GTCCTTGTTC TTATTTGGAG
1101 AAGTTCACAA AGCGCTCTGG AAGACGGAGC AGGGGACTGT CGTAGGGATC
1151 CTCAATGCCA ACCCCATGAA GCCCAAGGAT GGTTCAGAGG AGGTGTGTTT
1201 ATCTATCGAT CATCCTCAGA AGGTCTTAAT TATGGGTGAA GCTCTTGACC
1251 TGGGAACCTG TAAAGCCAAG AAGAAGAATG GAGAGCCGTG CACGCAGACT
1301 GTGAATTTGC GTGACTGTGA GTACTGTCAG TACCATGTCC AGGCTCAGTA
1351 CAAGAAGCTC AGTGCAAAGC GTGCGGATCT GCAGTCCACC TTCTCTGGAG
1401 GACGAATTCC AAAGAAGTTT GCCCGCAGAG GCACCAGCCT CAAAGAACGG
1451 CTGTGCCAAG ATGGCTTTTA CTACGGAGGG GTTTCTTCTG CCTCGTATGC
1501 AGCTTCAATT GCAGCAGCTG TGGCTCCTAA GAAGAAGATT CAAACCACTC
1551 TGAGTAATCT GGTTGTTAAG GGCACAAACT TGATCATCCA GGAAACACGG
1601 CAAAAACTCG GAATACCCCA GAAGAGCCTG TCTTGCTCTG AGGAGTTCAA
1651 GGAACTGATG GACCTGCCGA CGTGTGGAGC CAGGAACTTA AAACAACATT
1701 TAGCCAAAGC CTCAGCTTCA GGGATTATGG GGAGCCCAAA ACCAGCCATC
1751 AAGTCCATCT CGGCCTCAGC ACTCTTGAAG CAACAGAAGC AGCGGATGTT
1801 GGAGATGAGG AGAAGGAAAT CAGAAGAAAT ACAGAAGCGA TTTCTGCAGA
1851 GCTCAAGTGA AGTTGAGAGC CCAGCTGTGC CATCTTCATC AAGACAGCCC
1901 CCTGCTCAGC CTCCACGGAC AGGATCCGAG TTCCCCAGGC TGGAGGGAGC
1951 CCCGGCCACA ATGACGCCCA AGCTGGGGCG AGGTGTCTTG GAAGGAGATG
2001 ATGTTCTCTT TTATGATGAG TCACCACCAC CAAGACCAAA ACTGAGTGCT
2051 TTAGCAGAAG CCAAAAAGTT AGCTGCTATC ACCAAATTAA GGGCAAAAGG
2101 CCAGGTTCTT ACAAAAAACA ACCCAAACAG CATTAAGAAG AAACAAAAGG
2151 ACCCTCAGGA CATCCTGGAG GTGAAGGAAC GTGTAGAAAA AAACACCATG
2201 TTTTCTTCTC AAGCTGAGGA TGAATTGGAG CCTGCCAGGA AAAAAAGGAG
2251 AGAACAACTT GCCTATCTGG AATCTGAGGA ATTTCAGAAA ATCCTAAAAG
2301 CAAAATCAAA ACACACAGGC ATCCTGAAAG AGGCCGAGGC TGAGATGCAG
2351 GAGCGCTACT TTGAGCCACT GGTGAAAAAA GAACAAATGG AAGAAAAGAT
2401 GAGAAACATC AGAGAAGTGA AGTGCCGTGT CGTGACATGC AAGACGTGCG
2451 CCTATACCCA CTTCAAGCTG CTGGAGACCT GCGTCAGTGA GCAGCATGAA
2501 TACCACTGGC ATGATGGTGT GAAGAGGTTT TTCAAATGTC CCTGTGGAAA
2551 CAGAAGCATC TCCTTGGACA GACTCCCGAA CAAGCACTGC AGTAACTGTG
2601 GCCTCTACAA ATGGGAACGG GACGGAATGC TAAAGGTATG CCATTTGCGT
2651 ACTAATTTTT GACTCCTTTT AGTGACCCAT GCTAATAATG TGGAACCATC
```

```
2701 TCCTATTAAA ATATTTTCAT TTTTCTAGGA AAAGACTGGT CCAAAGATAG
2751 GAGGAGAAAC TCTGTTACCA AGAGGAGAAG AACATGCTAA ATTTCTGAAC
2801 AGCCTTAAAT AACCCGAACT TCAGACATTT TCCCACAGAC TTCCTGGCCT
2851 CCTGTGACTC TGGAAAGCAA AGGATTGGCT GTGTATTGTC CATTGATTCC
2901 TGATTGACGC CGTCAAAAAC AAATGCTTGT TAAGCCCATA AGCTTTGCCT
2951 GCTTACTTTC TGCCATTGGG TTGGTTTGAT ACCACATTTA ACATTGACAT
3001 TTAAGTGGAA AACCAAGTTA TCATTGTCTT TTCTAAGCTC AGTGTGGATG
3051 ATTGCATTAC TTCATTCACT GAAGTTTTTG CCCAAAAATT GGAAGGTAAA
3101 CAGAGAGCTA TGTTTCTGTA TCTTTTGGTT ATAGAGTGTT CACTTCTTTA
3151 TCATAACAAA ATTCTAGTGT TTATACGAAC ACCCAGAGGC AAAAGAATTT
3201 GGCTTAATTC TCACTCCAGG TAAGTAGCTT AACTTCTGGG CTTCAGTTTT
3251 CTCATCTGTA AAATCAGGAA GATTGGACTA AGTGATCCTG AAATGTATTT
3301 TTTAGCACTG GATTTCTACA AATAATAAAA CTTTCCCATC TAGATAATGA
3351 TGATCACATA GTCTTGATGT ACGGACATTA AAAGCCAGAT TTCTTCATTC
3401 AATTCTGTTA TCTCTGTTTT ACTCTTTGAA ATTGATCAAG CCACTGAATC
3451 ACTTTGCATT TCAGTTTATA TATAGAGAGA GAAAGAAGGC TGTCTGCTCT
3501 TACATTATTG TGGAGCCCTG TGATAGAAAT ATGTAAAATC TCATATTATT
3551 TTTTTTTTAA TTTTTTTATT TTTTATGACA GGGTCTCACT ATGTCACCCT
3601 GGCTGGAGTG CAGTAGTGCG ATCGCGGCAC ACTGCAGCCT TGGCTTCCCT
3651 GGGCTCAAGC AGTCCTCCCA CCTCAGTCTC CCAAATAGCT AGGACTACAG
3701 GCGTGCGTGA CCAAGCCCAG CTAATTTTTG CATTTTTTGT AGAGATGGGG
3751 TTTTGCCATG TTGCTCAGGC TGGTCTCAAA CTCCTGAGCA CTAGCAATCC
3801 ACCCACCTCT GTTTCCAAAA AAAAAAAAAA AATGAAAGGT CAACCCCTAT
3851 GCAAATTACC ACAGCAAAGG TTTCATTCAG GAGATTCTTC CATCTGGGCA
3901 ACCTGGTTTT CCAAATATCA TTTGACCTAA GTGAATGTTG ATACTAGCTA
3951 AAGATTGGGT AAATTGGTTG AATTATTGTA TTGAAGCTTG AGCTGTAGCT
4001 AAAAGTAATT TAGGTTTCCC CTAAGATGTT ATTATGTTAG GGACATAACA
4051 CTTTTGGGAG GTTGTTGTGG GAGATGGTTG ATTTAGGTTT TCAAAAGCTA
4101 GAAATAAAAT TTACATGCCT TAGATTTCAT AAAATTCTGC TCTAATTGGG
4151 TGGAAGGTGC TGTATCTAAC TTGTGTTCCT CCTAAGGTTA TGTCCTAATA
4201 ACTATTCTTT TAGGAGTATA CTTCTACTTT ATAGAAGGTT GCTTTTCTTT
4251 TTAATTTTTT CTAACAAAGA AAAGAATAAA GTATTTATTA ATAAGAACCA
4301 GAAAGCACTT GAAACTGATG TTTTTAATGG CTCATTTAGG GTAGATTTAT
4351 TTATCTCATT AACTTAAAAC AGCTATGTGT ATGAAATAGG TCACAACAGA
4401 ACTTGAACAC CAGGTTGGTG TCTGAGCAAT CCCTTTCTTA TGGGAAAAAC
4451 AATGTTCTTG TTTGAACAGA GGGTATCATT GCAGTCAGTA TTCACGTGTA
4501 TATTGTTATA TAAGTTGTAT AATATGCTTG TAAAGGCTGA GGGTGAGCTG
4551 TATCTGGATG CCTTTTTACA ATTTGATTTT AACTTTTAAA ATAAATTTAA
4601 AACATAAAAA AAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 95 bp to 2659 bp; peptide length: 855
Category: similarity to known protein
Classification: Cell division

```
  1 MDEEEDNLSL LTALLEENES ALDCNSEENN FLTRENGEPD AFDELFDADG
 51 DGESYTEEAD DGETGETRDE KENLATLFGD MEDLTDEEEV PASQSTENRV
101 LPAPAPRREK TNEELQEELR NLQEQMKALQ EQLKVTTIKQ TASPARLQKS
151 PEKSPRPPLK ERRVQRIQES TCFSAELDVP ALPRTKRVAR TPKPSPPDPK
201 SSSSRMTSAP SQPLQTISRN KPSGITRGQI VGTPGSSGET TQPICVEAFS
251 GLRLRRPRVS STEMNKKMTG RKLIRLSQIK EKMAREKLEE IDWVTFGVIL
301 KKVTPQSVNS GKTFSIWKLN DLRDLTQCVS LFLFGEVHKA LWKTEQGTVV
351 GILNANPMKP KDGSEEVCLS IDHPQKVLIM GEALDLGTCK AKKKNGEPCT
401 QTVNLRDCEY CQYHVQAQYK KLSAKRADLQ STFSGGRIPK KFARRGTSLK
451 ERLCQDGFYY GGVSSASYAA SIAAAVAPKK KIQTTLSNLV VKGTNLIIQE
501 TRQKLGIPQK SLSCSEEFKE LMDLPTCGAR NLKQHLAKAS ASGIMGSPKP
551 AIKSISASAL LKQQKQRMLE MRRRKSEEIQ KRFLQSSSEV ESPAVPSSSR
601 QPPAQPPRTG SEFPRLEGAP ATMTPKLGRG VLEGDDVLFY DESPPPRPKL
651 SALAEAKKLA AITKLRAKGQ VLTKTNPNSI KKKQKDPQDI LEVKERVEKN
701 TMFSSQAEDE LEPARKKRRE QLAYLESEEF QKILKAKSKH TGILKEAEAE
751 MQERYFEPLV KKEQMEEKMR NIREVKCRVV TCKTCAYTHF KLLETCVSEQ
801 HEYHWHDGVK RFFKCPCGNR SISLDRLPNK HCSNCGLYKW ERDGMLKVCH
851 LRTNF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2h15, frame 2

TREMBLNEW:SPBC1347_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division cycle protein 23"; S.pombe chromosome II cosmid c1347., N = 2, Score = 284, P = 7e-21

PIR:S48384 DNA43 protein - yeast (Saccharomyces cerevisiae), N = 2, Score = 203, P = 7e-12

TREMBL:SCDNA52A_1 gene: "DNA52"; Saccharomyces cerevisiae DNA52 gene, complete cds., N = 2, Score = 201, P = 7.9e-12

TREMBLNEW:AC006234_6 gene: "F5H14.6"; Arabidopsis thaliana chromosome II BAC F5H14 genomic sequence, complete sequence., N = 2, Score = 211, P = 1.7e-15

PIR:S48384 DNA43 protein - yeast (Saccharomyces cerevisiae), N = 2, Score = 203, P = 7.2e-12

```
>TREMBLNEW:SPBC1347_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division
     cycle protein 23"; S.pombe chromosome II cosmid c1347.
            Length = 593

  HSPs:

 Score = 284 (42.6 bits), Expect = 7.0e-21, Sum P(2) = 7.0e-21
 Identities = 97/383 (25%), Positives = 186/383 (48%)

Query:   109 EKTNEELQEELRNLQEQMKALQEQLKVTTIKQTASPARLQKSPEKSPRPPLKERRVQRIQ 168
             E+ + +L+E  + LQ Q+  +QE+ ++ + ++ AS    +    + PR P ++ RV  +
Sbjct:     8 EENDLDLEE--KRLQRQLNEIQEKKRLRSAQKEASSENAEVI--QVPRSPPQQVRVLTVS 63

Query:   169 ESTCFSAE----LDVPALPRTKRVARTPKPSPPDPKSSSSRMTSAPSQP------LQTIS 218
               +   +      L +       K V+      P  P PK    R+  A +Q       L+T+
Sbjct:    64 SPSKLKSPKRLILGIDKGKTGKDVSLGKGPRGPLPKPFHERLAEARNQERKRSDKLKTMK 123

Query:   219 RNKPSGITRGQIVGTPGSSGETTQPI-C--VEAFSGLRLRRPRVSSTEMNKKMTGRKLIR 275
             +N+      R + +   G S E   P+ C  ++ +S       +S          + G ++
Sbjct:   124 KNRKQSFQRKRNILEDGKSEEEKFPMKCDEIDPYSRQAIVIRYISDEVAKENIGGNQVYL 183

Query:   276 LSQIKEKMAREKLE--EID-WVTFGVILKKV-TPQSVNSGKTFSIWKLNDLRDLTQCVSL 331
             + Q+ + +    K E  E+D +V  G++      T ++VN  K + +  L DL+   +C
Sbjct:   184 IHQLLKLVRAPKFEAPEVDNYVVMGIVASNSGTRETVNGNK-YCMLTLTDLKWQLEC--- 239

Query:   332 FLFGEVHKALWKTEQGTVVGILNANPMKPKDGS-EEVCLSIDHPQKVLI-MGEALDLGTC 389
             FLFG+  +  WK + GTV+ +LN    +KPK+       L +D     VL+ +G +  LG C
Sbjct:   240 FLFGKAFERYWKIQSGTVIALLNPEVLKPKNPDIGRFSLKLDSEYDVLLEIGRSKHLGYC 299

Query:   390 KAKKKNGEPCTQTVNLRDCEYCQYHVQAQYKKLSAKRADLQSTFSGGRIPKKFARRGTSL 449
              +++K+GE C    ++ R  + C+YHV    ++  + R +  S+ +    P+  ARR
Sbjct:   300 SSRRKSGELCKHWLDKRAGDVCEYHVDLAVQRSMSTRTEFASSMATMHEPR--ARR---- 353

Query:   450 KERLCQDGF--YYGGVSSASYAASIAAAVAPKKKIQT 484
             ++R    GF  Y+ G   ++   ++A    +   +QT
Sbjct:   354 EKRFRGQGFQGYFAGEKYSAIPNAVAGLYDAEDAVQT 390

 Score = 41 (6.2 bits), Expect = 7.0e-21, Sum P(2) = 7.0e-21
 Identities = 12/43 (27%), Positives = 17/43 (39%)

Query:   453 LCQDGFYYGGVSSASYAASIAAAVAPKKKIQTTLSNLVVKGTN 495
             L +D           S  AS  A++     K       + SN  + GTN
Sbjct:   465 LSKDSEIDSSTKKPSVLASFNASIMNPKSSLPSFSNSAILGTN 507

 Score = 40 (6.0 bits), Expect = 8.9e-21, Sum P(2) = 8.9e-21
 Identities = 13/26 (50%), Positives = 18/26 (69%)

Query:   536 LAKASASGIMGSPKPAIKSISASALL 561
             LA  +AS IM +PK ++ S S SA+L
Sbjct:   481 LASFNAS-IM-NPKSSLPSFSNSAIL 504
```

Pedant information for DKFZphtes3_2h15, frame 2

Report for DKFZphtes3_2h15.2

```
[LENGTH]        855
[MW]            96135.01
[pI]            8.96
[HOMOL]         TREMBLNEW:SPBC1347_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division
cycle protein 23"; S.pombe chromosome II cosmid c1347. 5e-16
[FUNCAT]        03.22 cell cycle control and mitosis  [S. cerevisiae, YIL150c] 1e-11
[FUNCAT]        03.16 dna synthesis and replication  [S. cerevisiae, YIL150c] 1e-11
[FUNCAT]        30.10 nuclear organization    [S. cerevisiae, YIL150c] 1e-11
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY    12.05 %
[KW]            COILED_COIL        4.21 %

SEQ     MDEEEDNLSLLTALLEENESALDCNSEENNFLTRENGEPDAFDELFDADGDGESYTEEAD
SEG     .......................................................xxxxx
PRD     cccchhhhhhhhhhhhhhcccccccccceeeecccccccccceeeecccccccceeeeec
COILS   ............................................................

SEQ     DGETGETRDEKENLATLFGDMEDLTDEEEVPASQSTENRVLPAPAPRREKTNEELQEELR
SEG     xxxxxxxxxxxx.......................................xxxxxxxxx
PRD     cccccccccccchhhhhhcccccccccceeeccccccccccccccccccchhhhhhhhhhh
COILS   ...............................................CCCCCCCCCCCCC

SEQ     NLQEQMKALQEQLKVTTIKQTASPARLQKSPEKSPRPPLKERRVQRIQESTCFSAELDVP
SEG     xxxxx.......................................................
PRD     hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccceeeeeccccccccccccc
COILS   CCCCCCCCCCCCCCCCCCC.........................................

SEQ     ALPRTKRVARTPKPSPPDPKSSSSRMTSAPSQPLQTISRNKPSGITRGQIVGTPGSSGET
SEG     ..........xxxxxxxxxxxxx.....................................
PRD     cccccceeeecccccccccccccccchhhhhhccccchhhhhhcccccceeeeeccccccc
COILS   ............................................................

SEQ     TQPICVEAFSGLRLRRPRVSSTEMNKKMTGRKLIRLSQIKEKMAREKLEEIDWVTFGVIL
SEG     ............................................................
PRD     cccccccccchhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeeeeee
COILS   ............................................................

SEQ     KKVTPQSVNSGKTFSIWKLNDLRDLTQCVSLFLFGEVHKALWKTEQGTVVGILNANPMKP
SEG     ............................................................
PRD     cccccccccccceeeeeeecccchhhhhheeeeecchhhhhhhccceeeeecccccccccc
COILS   ............................................................

SEQ     KDGSEEVCLSIDHPQKVLIMGEALDLGTCKAKKKNGEPCTQTVNLRDCEYCQYHVQAQYK
SEG     ............................................................
PRD     ccccceeeeecccccceeeccccccccccccccccccccccceeeccccccchhhhhhhh
COILS   ............................................................

SEQ     KLSAKRADLQSTFSGGRIPKKFARRGTSLKERLCQDGFYYGGVSSASYAASIAAAVAPKK
SEG     .....................................xxxxxxxxxxxxxxxxxxxx...
PRD     hhhhhhhhhhhhcccccccccccccchhhhhhhcccccccccccchhhhhhhhhhhhcch
COILS   ............................................................

SEQ     KIQTTLSNLVVKGTNLIIQETRQKLGIPQKSLSCSEEFKELMDLPTCGARNLKQHLAKAS
SEG     ............................................................
PRD     hhhhhheeecccccceeeehhhhhhhcccccccchhhhhhhhhcccccchhhhhhhhhhhh
COILS   ............................................................

SEQ     ASGIMGSPKPAIKSISASALLKQQKQRMLEMRRRKSEEIQKRFLQSSSEVESPAVPSSSR
SEG     .............................................xxxxxxxxxxxxxxx
PRD     hhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccc
COILS   ............................................................

SEQ     QPPAQPPRTGSEFPRLEGAPATMTPKLGRGVLEGDDVLFYDESPPPRPKLSALAEAKKLA
SEG     xxxxxxxx........................................xxxxxxxxxxxx
PRD     ccccccccccccccccccccccccccccccccccccceeeeecccccchhhhhhhhhhhh
COILS   ............................................................

SEQ     AITKLRAKGQVLTKTNPNSIKKKQKDPQDILEVKERVEKNTMFSSQAEDELEPARKKRRE
SEG     xxxxx.......................................................
PRD     hhhhhhhheeeeecccccccccccccccchhhhhhhhhhhhcchhhhhhhhhhhhhhhhhh
COILS   ............................................................

SEQ     QLAYLESEEFQKILKAKSKHTGILKEAEAEMQERYFEPLVKKEQMEEKMRNIREVKCRVV
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheee
COILS   ............................................................

SEQ     TCKTCAYTHFKLLETCVSEQHEYHWHDGVKRFFKCPCGNRSISLDRLPNKHCSNCGLYKW
```

```
SEG     ............................................................
PRD     eeecceeeeeeecccceeeccccccccccceeeeeeccccccccccccccccccccceeec
COILS   ............................................................

SEQ     ERDGMLKVCHLRTNF
SEG     ...............
PRD     ccccccccccccccc
COILS   ...............
```

(No Prosite data available for DKFZphtes3_2h15.2)

(No Pfam data available for DKFZphtes3_2h15.2)

DKFZphtes3_2i5
----

group: testes derived

DKFZphtes3_2i5 encodes a novel 151 amino acid protein with weak similarity to. C.elegans cosmid F20D12.3

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans F20D12.3

many ATGs in front of the start of the ORF,
unspliced intron in 5' region?

Sequenced by EMBL

Locus: unknown

Insert length: 2142 bp
Poly A stretch at pos. 2121, polyadenylation signal at pos. 2102

```
   1 GCAGTAAATA TGATATGAAA GAATTCTCTA ACTTGGGGGT GGCTTGTAAC
  51 CTGTAATAAA AATATTGCTA AAATACCTTC TCTCACTTTG AAAAAGCATC
 101 TGAGCAATCC TCAGTTATTG GTGAATTCTT ACCAGTGTTT AATTCCTCTC
 151 TTTCCGTTAT GGTCTTAGTG TGGTTGTCCT GGTGTAGTAT TTCAAGAGGA
 201 ACCTGCAGCA AGATGAAAAG AGAGTGGGAC TTGGAGCTAA GAACGTTTTT
 251 GGCTTTAAGT GCTACGTTAA CTCATTAAAT TCTTAGTGAT CTTGGGGAAG
 301 TCCCCTCACC AGTGTGAGCC TCAGTTTTCT TATCTAATAA GTAAGGATAA
 351 TCTTACCCAC CTTATTGCGG GGGCCCGAGG ATTACATGAT TGGTGTAACA
 401 GTAGCACCTT GTACATTTGA AAGGACTAAT ACCAGTGGAC TTTAACCTTG
 451 GCTGGGCTTT GGAATTCTTG GTGGGACTTT TTAATCATGT AGATTCTCAG
 501 GCCCCTGCCT GGCCTGTGGA ACCACAGACT CTATAGGTGG GCCCTTCCAG
 551 AAGGCCTCAT GGGTGGTTCT CATGTGGAAC CTGTGTTGCA AGCCACTGCA
 601 TGGTGTTACT GCTATTAACA TTAAAACTTA TATTTTCCTT ATTGTGTGGA
 651 TATATCTGTG GTGTTTGCCC ATGTATACTT CATTTTACAT TTCTTAAAGA
 701 ATAGAATGGA ATGGTTTTAA GCACGCTACA TTGTCCAGGT TATACCCACA
 751 GAAGAGCTGT TGTGTAACAG AATCAGCATC ATACCTGAAT CATTTGTACA
 801 TTGCATATAA GACTATGTCT AAGTAGAAGA TGCTATGAAA TCATGTCTGC
 851 TGTGGGGCCA GGCATAATTA TGAATGTTAC TTAAGAGCAT AGGTGAGGTG
 901 AGAAAAGGGA ATGTGACTAG TGTTTTAGTA TTTTCTTGGT GTGGGATGAA
 951 GTATAATTCT TTTTTTTTTT TCTCAACAAA GCAGTAAAAC TAGAAAGAAG
1001 GAGAACTCTT CCCTCAAGAA TGGCTGTACC TTCATATCTA GAGGCACATT
1051 AAAAAAAAGA ACGTCTGTAC CTTAAAAATG GAGGTCATTT CATTGTGTTC
1101 ATTTTCAAGG TTGTTGTATG GCTCGGTCAG AACTTTCTGT TACCAGAAGA
1151 CACTCACATT CAGAATGCTC CATTTCAAGT GTGTTTCACA TCTTTACGGA
1201 ATGGCGGCCA CCTGCATATA AAAATAAAAC TTAGTGGAGA GATCACTATA
1251 AATACTGATG ATATTGATTT GGCTGGTGAT ATCATCCAGT CAATGGCATC
1301 ATTTTTTGCT ATTGAAGACC TTCAAGTAGA AGCGGATTTT CCTGTCTATT
1351 TTGAGGAATT ACGAAAGGTG CTAGTTAAGG TGGATGAATA TCATTCAGTG
1401 CATCAGAAGC TCAGTGCTGA TATGGCTGAT CATTCTAATT TGATCCGAAG
1451 TTTGCTGGTC GGAGCTGAGG ATGCTCGTCT GATGAGGGAC ATGAAAACAA
1501 TGAAGAGTCG TTATATGGAA CTCTATGACC TTAATAGAGA CTTGCTAAAT
1551 GGATATAAAA TTCGCTGTAA CAATCACACA GAGCTGTTGG GAAACCTCAA
1601 AGCAGTAAAT CAAGCAATTC AAAGAGCAGG TCGTCTGCGG GTTGGAAAAC
1651 CAAAGAACCA GGTGATCACT GCTTGTCGGG ATGCAATTCG AAGCAATAAC
1701 ATCAACACAC TGTTCAAAAT CATGCGAGTG GGGACAGCTT CTTCCTAGGT
1751 GAGGAAAATA CAGGTCATGA AGTTCCTGGC AAAGATTTTC TGTTAAAAAC
1801 CTATGCTGGT TTGCTTTGGA TCACACCCTG GTGAACCCCG GGTGCTAAGA
1851 ATGAAAATAA CCTTGGTGAG TTGTACAAAT TAAAGACAAA GAACTACATG
1901 TGAAGATAGA CTTGCTTTCT ATTTTTAAAT CAGTAGTAGT ACTGTTGCTG
1951 AATAATACTA GGTTTTTATG GAATAGGATG AATGCTTTTG AAGTATTAGG
2001 GCTTCAGAGT CCAATTTTGC TTATTTATGG TATATAAATA CATATTTTTT
2051 TCTTGAAATT GCAATTGAGT TTGTACTTTT CAAATAGATT ATCTACTTTT
2101 TCATTAAAAT GTAAAGATGT TAAAAAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

```
No Medline entry


                    Peptide information for frame 3
                    -------------------------------

ORF from 1293 bp to 1745 bp; peptide length: 151
Category: similarity to unknown protein
Classification: no clue

    1 MASFFAIEDL QVEÀDFPVYF EELRKVLVKV DEYHSVHQKL SADMADHSNL
   51 IRSLLVGAED ARLMRDMKTM KSRYMELYDL NRDLLNGYKI RCNNHTELLG
  101 NLKAVNQAIQ RAGRLRVGKP KNQVITACRD AIRSNNINTL FKIMRVGTAS
  151 S


                              BLASTP hits

No BLASTP hits available

             Alert BLASTP hits for DKFZphtes3_2i5, frame 3

TREMBL:CEF20D12_1 gene: "F20D12.3"; Caenorhabditis elegans cosmid
F20D12., N = 1, Score = 173, P = 4.5e-12


>TREMBL:CEF20D12_1 gene: "F20D12.3"; Caenorhabditis elegans cosmid F20D12.
            Length = 699

  HSPs:

 Score = 173 (26.0 bits), Expect = 4.5e-12, P = 4.5e-12
 Identities = 33/130 (25%), Positives = 72/130 (55%)

Query:     20 FEELRKVLVKVDEYHSVHQKLSADMADHSNLIRSLLVGAEDARLMRDMKTMKSRYMELYD 79
              F+E  ++L ++D     V  +L+A++ +     ++ +++ AED+  + ++    +  Y+ L
Sbjct:    569 FKEADEILEEIDPMTEVRDRLTAELQERQAAVKEIIIRAEDSIAIDNIPDARKFYIRLKA 628

Query:     80 LNRDLLNGYKIRCNNHTELLGNLKAVNQAIQRAGRLRVGKPKNQVITACRDAIRSNNINT 139
                +         ++R NN     + +L+ +N+ I+    RLRVG+P  Q++ +CR AI   +N
Sbjct:    629 NDAAARQAAQLRWNNQERCVKSLRRLNKIIENCSRLRVGEPGRQIVVSCRSAIADDNKQI 688

Query:    140 LFKIMRVGTA 149
              + KI++ G +
Sbjct:    689 ITKILQYGAS 698


             Pedant information for DKFZphtes3_2i5, frame 3
             ---------------------------------------------

                      Report for DKFZphtes3_2i5.3


[LENGTH]        151
[MW]            17304.07
[pI]            9.33
[HOMOL]         TREMBL:CEF20D12_1 gene: "F20D12.3"; Caenorhabditis elegans cosmid F20D12. 2e-12

[KW]            Alpha_Beta


SEQ     MASFFAIEDLQVEADFPVYFEELRKVLVKVDEYHSVHQKLSADMADHSNLIRSLLVGAED
PRD     ccceeeehhhhhhccccchhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ     ARLMRDMKTMKSRYMELYDLNRDLLNGYKIRCNNHTELLGNLKAVNQAIQRAGRLRVGKP
PRD     hhhhhcccchhhhhheeeccchhhhhhheeeeecccchhhhhhhhhhhhhhhhhcccccccc

SEQ     KNQVITACRDAIRSNNINTLFKIMRVGTASS
PRD     cceeeeehhhhhhcccceeeecccceeecccc


(No Prosite data available for DKFZphtes3_2i5.3)

(No Pfam data available for DKFZphtes3_2i5.3)
```

DKFZphtes3_2119
----

group: testes derived

DKFZphtes3_2119 encodes a novel 166 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, no EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1079 bp
Poly A stretch at pos. 1053, polyadenylation signal at pos. 1038

```
   1 CCACAGGACA CACTGTTCCC AGGGCACAGA CACCCTGGGC TTTGGTTGGG
  51 TCTTGGCCTC CAGGTAGGGC CCTGTTGGGC AGCGGGCAGC AACTCCTGAG
 101 ACACTACTGT GATTCTTGGT GGTGGCTGTG GTAAAAAACC TGCAGGGCTA
 151 GAGTTTGGGG TGAGATTCAG CAGTAACTGT GGCCTCTCCT AGTGACAGTA
 201 TGTCACTCCC ACTCCCAGCA CGCATGCCCA CAGGCCACGG CCTCCACATC
 251 ACAAACCCCC CACCAAGTTG CCCATCTATG GAGCAGCTCC CATACGGCAG
 301 GGTCAGGCTC TTACCTCCAC CTCCAGGGCA CAGACAGGGG GAGCTCTGTC
 351 TCACTGTAAG GCAATGAGGA GAGTTGAGGG CCCAGACCAG GCTAGGGGCC
 401 ATCCCCTTTC CCGAGCAGGC CTCAGGGAAG GACCAGCCCC ATTCCCATCT
 451 GACCTAGGTC TTAGCCCAGG AGCCTGCATA GGGAAGAAAG GACAGACAGG
 501 GCCTCCTTAC TGGCTGACAC TCAGGAGGGG CTGGGGCAAG AGAGCAGAGG
 551 GAGCGCAGGG CCAGGCAGGG GCTGCTGAGG ATCCATGGGA GCTCAGGGTG
 601 CACAAGGGGG CTGCCCTTCC TGGGCTGCAG GCAGCATCCC TATGGGAGCT
 651 GAGAAAGTCC AATCCTGAGA TGGGACAGTG CTGCCCAGGG GTGTGTGGCT
 701 GGGCCCTGAC AACAGTCTCC CCAAAAGTGA CCACATCACC AGGCTCAGTT
 751 CCAGGAAGGC TGAGAAGTGC CCAGTACACT GAGGATGCAC CTCAGTTACA
 801 TAAAATAAAT GAAACTGGAG TACTAACGTA CAGTTTAAAG GTTATAGTTA
 851 CTATTTTTAT ATGATATACT AGTAATTTTT GAATAGGGTA AACTTTAGGT
 901 GTTTTGACAC CAAAAGAAAA CTACATGAGT TCATGCATGT GTTAAATTGC
 951 TTTACTGTAG TAATCATTTA CATGTATATG TATATATGAA TATAATTATG
1001 GGCTCATTAA ATTTAAATAT TATAAATAGG TGACAAAGAA TAAAGTTAAC
1051 TGGAAAAAAA AAAAAAAAAA AAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 364 bp to 861 bp; peptide length: 166
Category: putative protein
Classification: no clue

```
   1 MRRVEGPDQA RGHPLSRAGL REGPAPFPSD LGLSPGACIG KKGQTGPPYW
  51 LTLRRGWGKR AEGAQGQAGA AEDPWELRVH KGAALPGLQA ASLWELRKSN
 101 PEMGQCCPGV CGWALTTVSP KVTTSPGSVP GRLRSAQYTE DAPQLHKINE
 151 TGVLTYSLKV IVTIFI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2119, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2119, frame 1

Report for DKFZphtes3_2119.1

```
[LENGTH]    166
[MW]        17691.35
[pI]        9.54
[KW]        All_Beta
[KW]        LOW_COMPLEXITY      7.23 %

SEQ   MRRVEGPDQARGHPLSRAGLREGPAPFPSDLGLSPGACIGKKGQTGPPYWLTLRRGWGKR
SEG   ............................................................
PRD   cccccccccccccccccccccccccccccccccccccceeeccccccccccceeeeeccccc

SEQ   AEGAQGQAGAAEDPWELRVHKGAALPGLQAASLWELRKSNPEMGQCCPGVCGWALTTVSP
SEG   xxxxxxxxxxxx................................................
PRD   ccccccccccccccceeeeccccccccchhhhhhhhhhccccccccccccccceeeeeccc

SEQ   KVTTSPGSVPGRLRSAQYTEDAPQLHKINETGVLTYSLKVIVTIFI
SEG   ..............................................
PRD   ccccccccccccccccccccccccceeecccccceeeehhhhhhccc
```

(No Prosite data available for DKFZphtes3_2119.1)

(No Pfam data available for DKFZphtes3_2119.1)

DKFZphtes3_2m18
----

group: nucleic acid management

DKFZphtes3_2m18 encodes a novel amino acid protein, with similarity to mouse Dhml.

The protein seems to play a role in nucleotide metabolism, RNA metabolism, but also in DNA repair and cell cycle. The yeast homologue is a DNA strand exchange protein required for sporulation and homologous recombination.

The novel protein can find application as multifunctional nuclease / exoribonuclease.

nearly identical to mouse Dhml

complete cDNA, complete cds, start at Bp 42, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3022 bp
Poly A stretch at pos. 3004, polyadenylation signal at pos. 2981

```
   1 CTCGTCAGCC GGTCGGCCGC CGCCTCCAGC CGTGTGCCGC TATGGGAGTC
  51 CCGGCGTTCT TCCGCTGGCT CAGCCGCAAG TACCCGTCCA TCATAGTCAA
 101 CTGCGTGGAA GAGAAGCCAA AAGAATGCAA TGGTGTAAAG ATTCCAGTTG
 151 ATGCCAGTAA ACCTAATCCA AATGATGTGG AGTTTGATAA TCTGTATTTG
 201 GATATGAATG GAATCATCCA TCCCTGTACT CATCCTGAAG ACAAACCAGC
 251 ACCAAAAAAT GAAGATGAAA TGATGGTTGC AATTTTTGAG TACATTGACA
 301 GACTTTTCAG TATTGTAAGA CCAAGAAGAC TTCTCTACAT GGCAATAGAT
 351 GGAGTGGCAC CACGTGCTAA AATGAACCAG CAGCGTTCAA GGAGGTTCAG
 401 GGCATCAAAA GAAGGAATGG AAGCAGCAGT CGAGAAGCAG CGAGTCAGGG
 451 AAGAAATATT GGCAAAAGGT GGCTTTCTTC CTCCAGAAGA AATAAAAGAA
 501 AGATTTGACA GCAACTGTAT TACACCAGGA ACTGAATTCA TGGACAATCT
 551 TGCTAAATGC CTTCGCTATT ACATAGCTGA TCGTTTAAAT AATGACCCTG
 601 GGTGGAAAAA TTTGACAGTT ATTTTATCTG ATGCTAGTGC TCCTGGTGAA
 651 GGAGAACATA AAATCATGGA TTACATTAGA AGGCAAAGAG CCCAGCCTAA
 701 CCATGACCCA AATACTCATC ATTGTTTATG TGGAGCAGAT GCTGATCTCA
 751 TTATGCTTGG CCTTGCCACA CATGAACCGA ACTTTACCAT TATTAGAGAA
 801 GAATTCAAAC CAAACAAGCC CAAACCATGT GGTCTTTGTA ATCAGTTTGG
 851 ACATGAGGTC AAAGATTGTG AAGGTTTGCC AAGAGAAAAG AAGGGAAAGC
 901 ATGATGAACT TGCCGATAGT CTTCCTTGTG CAGAAGGAGA GTTTATCTTC
 951 CTTCGGCTTA ATGTTCTTCG TGAGTATTTG GAAAGAGAAC TCACAATGGC
1001 CAGCCTACCA TTCACATTTG ATGTTGAGAG GAGCATTGAT GACTGGGTTT
1051 TCATGTGCTT CTTTGTGGGA AATGACTTCC TCCCTCATTT GCCATCGTTA
1101 GAGATTAGGG AAAATGCAAT TGACCGTTTG GTTAACATAT ACAAAAATGT
1151 GGTACACAAA ACTGGGGGTT ACCTTACAGA AAGTGGTTAT GTCAATCTGC
1201 AAAGAGTACA GATGATCATG TTAGCAGTTG GTGAAGTTGA GGATAGCATT
1251 TTTAAAAAGA GAAAGGATGA TGAGGACAGT TTTAGAAGAC GACAGAAAGA
1301 AAAAAGAAAG AGAATGAAGA GAGATCAACC AGCTTTCACT CCTAGTGGAA
1351 TATTAACTCC TCATGCCTTG GGTTCAAGAA ATTCACCAGG TTCTCAAGTA
1401 GCCAGTAATC CGAGACAAGC AGCCTATGAA ATGAGGATGC AGAATAACTC
1451 TAGTCCTTCG ATATCTCCTA ATACGAGTTT CACATCTGAT GGCTCCCCGT
1501 CTCCATTAGG AGGAATTAAG CGAAAAGCAG AAGACAGTGA CAGTGAACCT
1551 GAGCCAGAGG ATAATGTCAG GTTATGGGAA GCTGGCTGGA AGCAGCGGTA
1601 CTACAAGAAC AAATTTGATG TGGATGCAGC TGATGAGAAA TTCCGTCGGA
1651 AAGTTGTGCA GTCGTACGTT GAAGGACTTT GCTGGGTTCT TAGATATTAT
1701 TACCAGGGCT GTGCTTCCTG GAAGTGGTAT TATCCATTTC ATTATGCACC
1751 ATTTGCTTCA GACTTTGAAG GCATTGCAGA CATGCCATCT GATTTTGAGA
1801 AGGGTACGAA ACCGTTTAAA CCACTAGAAC AACTTATGGG GGTATTTCCA
1851 GCTGCAAGTG GTAATTTTCT ACCTCCATCA TGGCGGAAGC TCATGAGTGA
1901 TCCTGATTCT AGTATAATTG ACTTCTATCC TGAAGATTTT GCTATTGATT
1951 TGAATGGGAA GAAATATGCA TGGCAAGGTG TTGCTCTCTT GCCATTCGTG
2001 GATGAGCGAA GGCTACGAGC TGCCCTAGAA GAGGTATACC CAGACCTCAC
2051 TCCAGAAGAG ACCAGAAGAA ACAGCCTTGG AGGTGATGTC TTATTTGTGG
2101 GGAAACATCA CCCACTCCAT GACTTCATTT TAGAGCTGTA CCAGACAGGT
2151 TCCACAGAGC CAGTGGAGGT ACCCCCTGAA CTATGTCATG GGATTCAAGG
2201 AAAGTTTTCT TTGGATGAAG AAGCCATTCT TCCAGATCAA ATAGTATGTT
2251 CTCCTGTTCC TATGTTAAGG GATCTGACAC AGAACACTGT AGTCAGTATT
2301 AATTTTAAAG ACCCACAGTT TGCTGAAGAT TACATTTTTA AAGCTGTAAT
2351 GCTTCCAGGA GCAAGAAAGC CAGCAGCAGT ACTGAAACCT AGTGACTGGG
2401 AAAAATCCAG CAATGGACGG CAGTGGAAGC CTCAGCTTGG CTTTAACCGT
2451 GACCGGAGGC CTGTGCACCT GGATCAGGCA GCCTTCAGGA CTTTGGGCCA
2501 TGTGATGCCA AGAGGCTCAG GAACTGGCAT TTACAGCAAT GCTGCACCAC
2551 CACCTGTGAC TTACCAGGGA AACTTATACA GGCCGCTTTT GAGAGGACAA
2601 GCCCAGATTC CAAAACTTAT GTCAAATATG AGGCCCCAGG ATTCCTGGCG
2651 AGGTCCTCCT CCCCTTTTCC AGCAGCAAAG GTTTGACAGA GGCGTTGGGG
```

```
2701 CTGAACCTCT GCTCCCATGG AACCGGATGC TGCAAACCCA GAATGCAGCC
2751 TTCCAGCCAA ACCAGTACCA GATGCTAGCT GGGCCTGGTG GGTATCCACC
2801 CAGACGAGAT GATCGTGGAG GGAGACAGGG ATATCCCAGA GAAGGAAGGA
2851 AATACCCTTT GCCACCACCC TCAGGAAGAT ACAATTGGAA TTAAGCTTTT
2901 GTAAAGCTTT CCCAAATCCT TTCATCATTC TACAGTTTTA TGCTATTTGT
2951 GGAAAGATTT CCTTCTCAAG TAGTAGTTTT TAATAAAACT ACAGTACTTT
3001 GTGTAAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

95192042:
Characterization of cDNA encoding mouse homolog of fission yeast dhp1+ gene: structural and functional conservation.

97361754:
Cloning and characterization of mouse Dhm2 cDNA, a functional homolog of budding yeast SEP1.

Peptide information for frame 3

ORF from 42 bp to 2891 bp; peptide length: 950
Category: strong similarity to known protein

```
  1 MGVPAFFRWL SRKYPSIIVN CVEEKPKECN GVKIPVDASK PNPNDVEFDN
 51 LYLDMNGIIH PCTHPEDKPA PKNEDEMMVA IFEYIDRLFS IVRPRRLLYM
101 AIDGVAPRAK MNQQRSRRFR ASKEGMEAAV EKQRVREEIL AKGGFLPPEE
151 IKERFDSNCI TPGTEFMDNL AKCLRYYIAD RLNNDPGWKN LTVILSDASA
201 PGEGEHKIMD YIRRQRAQPN HDPNTHHCLC GADADLIMLG LATHEPNFTI
251 IREEFKPNKP KPCGLCNQFG HEVKDCEGLP REKKGKHDEL ADSLPCAEGE
301 FIFLRLNVLR EYLERELTMA SLPFTFDVER SIDDWVFMCF FVGNDFLPHL
351 PSLEIRENAI DRLVNIYKNV VHKTGGYLTE SGYVNLQRVQ MIMLAVGEVE
401 DSIFKKRKDD EDSFRRRQKE KRKRMKRDQP AFTPSGILTP HALGSRNSPG
451 SQVASNPRQA AYEMRMQNNS SPSISPNTSF TSDGSPSPLG GIKRKAEDSD
501 SEPEPEDNVR LWEAGWKQRY YKNKFDVDAA DEKFRRKVVQ SYVEGLCWVL
551 RYYYQGCASW KWYYPFHYAP FASDFEGIAD MPSDFEKGTK PFKPLEQLMG
601 VFPAASGNFL PPSWRKLMSD PDSSIIDFYP EDFAIDLNGK KYAWQGVALL
651 PFVDERRLRA ALEEVYPDLT PEETRRNSLG GDVLFVGKHH PLHDFILELY
701 QTGSTEPVEV PPELCHGIQG KFSLDEEAIL PDQIVCSPVP MLRDLTQNTV
751 VSINFKDPQF AEDYIFKAVM LPGARKPAAV LKPSDWEKSS NGRQWKPQLG
801 FNRDRRPVHL DQAAFRTLGH VMPRGSGTGI YSNAAPPPVT YQGNLYRPLL
851 RGQAQIPKLM SNMRPQDSWR GPPPLFQQQR FDRGVGAEPL LPWNRMLQTQ
901 NAAFQPNQYQ MLAGPGGYPP RRDDRGGRQG YPREGRKYPL PPPSGRYNWN
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2m18, frame 3

PIR:I49635 mouse Dhm1 protein - mouse, N = 1, Score = 4765, P = 0

PIR:S43891 dhp1 protein - fission yeast (Schizosaccharomyces pombe), N = 3, Score = 1172, P = 2e-197

PIR:S20126 exoribonuclease RAT1 (EC 3.1.11.-) - yeast (Saccharomyces cerevisiae), N = 2, Score = 1146, P = 3.8e-175

PIR:S72531 exonuclease II - fission yeast (Schizosaccharomyces pombe), N = 4, Score = 622, P = 4.2e-125

>PIR:I49635 mouse Dhm1 protein - mouse
Length = 947

HSPs:

```
Score = 4765 (714.9 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 884/930 (95%), Positives = 895/930 (96%)

Query:     1 MGVPAFFRWLSRKYPSIIVNCVEEKPKECNGVKIPVDASKPNPNDVEFDNLYLDMNGIIH 60
             MGVPAFFRWLSRKYPSIIVNCVEEKPKECNGVKIPVDASKPNPNDVEFDNLYLDMNGIIH
Sbjct:     1 MGVPAFFRWLSRKYPSIIVNCVEEKPKECNGVKIPVDASKPNPNDVEFDNLYLDMNGIIH 60

Query:    61 PCTHPEDKPAPKNEDEMMVAIFEYIDRLFSIVRPRRLLYMAIDGVAPRAKMNQQRSRRFR 120
             PCTHPEDKPAPKNEDEMMVAIFEYIDRLF+IVRPRRLLYMAIDGVAPRAKMNQQRSRRFR
Sbjct:    61 PCTHPEDKPAPKNEDEMMVAIFEYIDRLFNIVRPRRLLYMAIDGVAPRAKMNQQRSRRFR 120

Query:   121 ASKEGMEAAVEKQRVREEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYYIAD 180
             A K GMEAAVEKQRVREEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYYIAD
Sbjct:   121 AIKGGMEAAVEKQRVREEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYYIAD 180

Query:   181 RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNHDPNTHHCLCGADADLIMLG 240
             RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPN DPNTHHCLCGADADLIMLG
Sbjct:   181 RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNQDPNTHHCLCGADADLIMLG 240

Query:   241 LATHEPNFTIIREEFKPNKPKPCGLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE 300
             LATHEPNFTIIREEFKPNKPKPC LCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE
Sbjct:   241 LATHEPNFTIIREEFKPNKPKPCALCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE 300

Query:   301 FIFLRLNVLREYLERELTMASLPFTFDVERSIDDWVFMCFFVGNDFLPHLPSLEIRENAI 360
             FIFLRLNVLREYLERELTMASLPF FDVERS DDW FMCFFVGNDFLPHLPSLEIRE AI
Sbjct:   301 FIFLRLNVLREYLERELTMASLPFPFDVERSNDDWEFMCFFVGNDFLPHLPSLEIREGAI 360

Query:   361 DRLVNIYKNVVHKTGGYLTESGYVNLQRVQMIMLAVGEVEDSIFKKRKDDEDSFRRRQKE 420
             DRLVNIYKNVVHKTGGYLTESGYVNLQRVQMIMLAVGEVEDSIFKKRKDDEDSFRRRQKE
Sbjct:   361 DRLVNIYKNVVHKTGGYLTESGYVNLQRVQMIMLAVGEVEDSIFKKRKDDEDSFRRRQKE 420

Query:   421 KRKRMKRDQPAFTPSGILTPHALGSRNSPGSQVASNPRQAAYEMRMQNNSSPSISPNTSF 480
             KRKRMKRDQPAFTPSGILTPHALGSRNSPG QVASNPRQAAYEMRMQ NSSPSISPNTSF
Sbjct:   421 KRKRMKRDQPAFTPSGILTPHALGSRNSPGCQVASNPRQAAYEMRMQRNSSPSISPNTSF 480

Query:   481 TSDGSPSPLGGIKRKAEDSDSEPEPEDNVRLWEAGWKQRYYKNKFDVDAADEKFRRKVVQ 540
              SDGSPSPLGGI+RKAEDSDSEPEPEDNVRLWEAGWKQRYYKNKFDVDAADEKFRRKVVQ
Sbjct:   481 ASDGSPSPLGGIRRKAEDSDSEPEPEDNVRLWEAGWKQRYYKNKFDVDAADEKFRRKVVQ 540

Query:   541 SYVEGLCWVLRYYYQGCASWKWYYPFHYAPFASDFEGIADMPSDFEKGTKPFKPLEQLMG 600
             SYVEGLCWVLRYYYQGCASWKW YPFHYAPFASDFEGIADM S+FEKGTKPFKPLEQLMG
Sbjct:   541 SYVEGLCWVLRYYYQGCASWKWLYPFHYAPFASDFEGIADMSSEFEKGTKPFKPLEQLMG 600

Query:   601 VFPAASGNFLPPSWRKLMSDPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA 660
             VFPAASGNFLPP+WRKLMSDPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA
Sbjct:   601 VFPAASGNFLPPTWRKLMSDPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA 660

Query:   661 ALEEVYPDLTPEETRRNSLGGDVLFVGKHHPLHDFILELYQTGSTEPVEVPPELCHGIQG 720
             ALEEVYPDLTPEE RRNSLGGDVLFVGK HPL DFILELYQTGSTEPV+VPPELCHGIQG
Sbjct:   661 ALEEVYPDLTPEENRRNSLGGDVLFVGKLHPLRDFILELYQTGSTEPVDVPPELCHGIQG 720

Query:   721 KFSLDEEAILPDQIVCSPVPMLRDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAAV 780
              FSLDEEAILPDQ VCSPVPMLRDLTQNT VSINFKDPQFAEDY+FKA MLPGARKPA V
Sbjct:   721 TFSLDEEAILPDQTVCSPVPMLRDLTQNTAVSINFKDPQFAEDYVFKAAMLPGARKPATV 780

Query:   781 LKPSDWEKSSNGRQWKPQLGFNRDRRPVHLDQAAFRTLGHVMPRGSGTGIYSNAAPPPVT 840
             LKP DWEKSSNGRQWKPQLGFNRDRRPVHLDQAAFRTLGHV PRGSGT +Y+N A  P
Sbjct:   781 LKPGDWEKSSNGRQWKPQLGFNRDRRPVHLDQAAFRTLGHVTPRGSGTSVYTNTALLPAN 840

Query:   841 YQGNLYRPLLRGQAQIPKLMSNMRPQDSWRGPPPLFQQQRFDRGVGAEPLLPWNRMLQTQ 900
             YQGN YRPLLRGQAQIPKLMSNMRP+DSWRGPPPLFQQ RF+R VGAEPLLPWNRM+Q Q
Sbjct:   841 YQGNNYRPLLRGQAQIPKLMSNMRPKDSWRGPPPLFQQHRFERSVGAEPLLPWNRMIQNQ 900

Query:   901 NAAFQPNQYQMLAGPGGYPPRRDD-RGGRQ 929
             NAAFQPNQYQML GPGGYPPRRDD RGGRQ
Sbjct:   901 NAAFQPNQYQMLGGPGGYPPRRDDHRGGRQ 930
```

Pedant information for DKFZphtes3_2m18, frame 3

Report for DKFZphtes3_2m18.3

[LENGTH]	950	
[MW]	108582.68	
[pI]	7.26	
[HOMOL]	PIR:I49635 mouse Dhm1 protein - mouse 0.0	
[FUNCAT]	08.01 nuclear transport	[S. cerevisiae, YOR048c] 1e-123
[FUNCAT]	04.01.04 rrna processing	[S. cerevisiae, YOR048c] 1e-123

```
[FUNCAT]      30.10 nuclear organization    [S. cerevisiae, YOR048c] 1e-123
[FUNCAT]      01.03.16 polynucleotide degradation   [S. cerevisiae, YGL173c] 3e-79
[FUNCAT]      30.03 organization of cytoplasm       [S. cerevisiae, YGL173c] 3e-79
[FUNCAT]      03.22 cell cycle control and mitosis  [S. cerevisiae, YGL173c] 3e-79
[PIRKW]       nucleus 1e-126
[PIRKW]       hydrolase 1e-122
[PIRKW]       exoribonuclease 1e-122
[PROSITE]     MYRISTYL         7
[PROSITE]     AMIDATION        2
[PROSITE]     CAMP_PHOSPHO_SITE      1
[PROSITE]     CK2_PHOSPHO_SITE       12
[PROSITE]     TYR_PHOSPHO_SITE       1
[PROSITE]     GLYCOSAMINOGLYCAN      1
[PROSITE]     PKC_PHOSPHO_SITE       8
[PROSITE]     ASN_GLYCOSYLATION      4
[KW]          TRANSMEMBRANE  1
[KW]          LOW_COMPLEXITY     6.21 %

SEQ   MGVPAFFRWLSRKYPSIIVNCVEEKPKECNGVKIPVDASKPNPNDVEFDNLYLDMNGIIH
SEG   ............................................................
PRD   cccchhhhhhhhcceeeeeeeecccccccccccccccccccccccccccceeeecccceee
MEM   ............................................................

SEQ   PCTHPEDKPAPKNEDEMMVAIFEYIDRLFSIVRPRRLLYMAIDGVAPRAKMNQQRSRRFR
SEG   ............................................................
PRD   ccccccccccccchhhhhhhhhhhhhhhhhcceeeeeeeeccccchhhhhhhhhhhhhhh
MEM   ............................................................

SEQ   ASKEGMEAAVEKQRVREEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYYIAD
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccchhhhhhhhhhhhhhh
MEM   ............................................................

SEQ   RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNHDPNTHHCLCGADADLIMLG
SEG   ............................................................
PRD   hcccccccceeeeeeeecccccccccchhhhhhhhhhhcccccccccccccccccccceeec
MEM   ............................................................

SEQ   LATHEPNFTIIREEFKPNKPKPCGLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE
SEG   ............................................................
PRD   cccccccccccccccccccccceeeccccccccccccccchhhhhhhhhcccccccccc
MEM   ............................................................

SEQ   FIFLRLNVLREYLERELTMASLPFTFDVERSIDDWVFMCFFVGNDFLPHLPSLEIRENAI
SEG   ............................................................
PRD   ccchhhhhhhhhhhhhhhhhchhhhhhhhhhhhheeeeeecccccccccccccccchhhh
MEM   ............................MMMMMMMMMMMMMMMMMMMM............

SEQ   DRLVNIYKNVVHKTGGYLTESGYVNLQRVQMIMLAVGEVEDSIFKKRKDDEDSFRRRQKE
SEG   ......................................................xxxxxx
PRD   hhhhhhhhhhccccccccccchhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhh
MEM   ............................................................

SEQ   KRKRMKRDQPAFTPSGILTPHALGSRNSPGSQVASNPRQAAYEMRMQNNSSPSISPNTSF
SEG   xxxxxxx........................................xxxxxxxxxxxxx
PRD   hhhhhhhcccccccccccccccccccccccchhhhhhhhhhhhhhhhcccccccccccccc
MEM   ............................................................

SEQ   TSDGSPSPLGGIKRKAEDSDSEPEPEDNVRLWEAGWKQRYYKNKFDVDAADEKFRRKVVQ
SEG   xx.............xxxxxxxxxxx..................................
PRD   ccccccccchhhhhhhcccccccchhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhh
MEM   ............................................................

SEQ   SYVEGLCWVLRYYYQGCASWKWYYPFHYAPFASDFEGIADMPSDFEKGTKPFKPLEQLMG
SEG   ............................................................
PRD   hhhhhheeeeeecccccccccccccccccccccccccccccccccccccccccchhhhhh
MEM   ............................................................

SEQ   VFPAASGNFLPPSWRKLMSDPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA
SEG   ............................................................
PRD   hcccccccccccccccccccccceeeccccceeeccccccceeeeeeeeeecccchhhhhh
MEM   ............................................................

SEQ   ALEEVYPDLTPEETRRNSLGGDVLFVGKHHPLHDFILELYQTGSTEPVEVPPELCHGIQG
SEG   ............................................................
PRD   hhhhhccccchhhhhhcccccceeeeeecccchhhhhhhhhcccccceeecccccccccc
MEM   ............................................................

SEQ   KFSLDEEAILPDQIVCSPVPMLRDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAAV
SEG   ............................................................
```

```
PRD   cccccceeecccceeeccccccccccccceeeeeccccccchhhhheeeccccccccccee
MEM   ............................................................

SEQ   LKPSDWEKSSNGRQWKPQLGFNRDRRPVHLDQAAFRTLGHVMPRGSGTGIYSNAAPPPVT
SEG   ............................................................
PRD   ecccccccccccccccccccccccccccchhhhhhhhhcccccccccccccccccccccc
MEM   ............................................................

SEQ   YQGNLYRPLLRGQAQIPKLMSNMRPQDSWRGPPPLFQQQRFDRGVGAEPLLPWNRMLQTQ
SEG   ............................................................
PRD   cccccchhhhcccchhhhhcccccccccccccccchhhhhcccccccccccccchhhhhh
MEM   ............................................................

SEQ   NAAFQPNQYQMLAGPGGYPPRRDDRGGRQGYPREGRKYPLPPPSGRYNWN
SEG   .............xxxxxxxxxxxxxxxxxxxx.................
PRD   hcccccccceeecccccccccccccccccccccccccccccccccccccc
MEM   ..................................................
```

Prosite for DKFZphtes3_2m18.3

PS00001	190->194	ASN_GLYCOSYLATION	PDOC00001
PS00001	247->251	ASN_GLYCOSYLATION	PDOC00001
PS00001	468->472	ASN_GLYCOSYLATION	PDOC00001
PS00001	477->481	ASN_GLYCOSYLATION	PDOC00001
PS00002	826->830	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	675->679	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	413->416	PKC_PHOSPHO_SITE	PDOC00005
PS00005	559->562	PKC_PHOSPHO_SITE	PDOC00005
PS00005	613->616	PKC_PHOSPHO_SITE	PDOC00005
PS00005	674->677	PKC_PHOSPHO_SITE	PDOC00005
PS00005	868->871	PKC_PHOSPHO_SITE	PDOC00005
PS00005	944->947	PKC_PHOSPHO_SITE	PDOC00005
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	331->335	CK2_PHOSPHO_SITE	PDOC00006
PS00006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS00006	501->505	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	583->587	CK2_PHOSPHO_SITE	PDOC00006
PS00006	619->623	CK2_PHOSPHO_SITE	PDOC00006
PS00006	624->628	CK2_PHOSPHO_SITE	PDOC00006
PS00006	670->674	CK2_PHOSPHO_SITE	PDOC00006
PS00006	723->727	CK2_PHOSPHO_SITE	PDOC00006
PS00006	784->788	CK2_PHOSPHO_SITE	PDOC00006
PS00007	659->667	TYR_PHOSPHO_SITE	PDOC00007
PS00008	125->131	MYRISTYL	PDOC00008
PS00008	375->381	MYRISTYL	PDOC00008
PS00008	450->456	MYRISTYL	PDOC00008
PS00008	600->606	MYRISTYL	PDOC00008
PS00008	825->831	MYRISTYL	PDOC00008
PS00008	829->835	MYRISTYL	PDOC00008
PS00008	926->932	MYRISTYL	PDOC00008
PS00009	638->642	AMIDATION	PDOC00009
PS00009	934->938	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_2m18.3)

DKFZphtes3_2m20
----

group: testes derived

DKFZphtes3_2m20 encodes a novel 183 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

group: unknown

DKFZphtes3_2m20 encodes a novel

amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

EST hits are only from testis or uterus librarys
remaining intron in3' UTR see EST-BLAST

Sequenced by EMBL

Locus: unknown

Insert length: 1341 bp
Poly A stretch at pos. 1320, polyadenylation signal at pos. 1300

```
   1 GCAATCCAGG AGCTGAATGG TAACTCTTCC ACAAGCGAAA ACTGTTCGTG
  51 AATACAAGCA AAAGGCCCCC CAAGAGGACC CCTGATATGA TCCAGCAGCC
 101 TCGGGCCCCG CTGGTGTTGG AGAAGGCTTC TGGTGAAGGA TTTGGCAAAA
 151 CCGCCGCTAT TATACAGCTC GCTCCTAAAG CTCCTGTTGA CCTGTGTGAG
 201 ACAGAGAAAC TGAGGGCAGC CTTCTTTGCA GTCCCGTTGG AAATGAGAGG
 251 GTCCTTCCTG GTGCTGCTCC TGAGGGAATG CTTCCGAGAC CTGAGCTGGC
 301 TGGCACTCAT CCATAGCGTC CGTGGGGAGG CGGGGCTGCT GGTGACGAGT
 351 ATCGTCCCGA AGACCCCGTT TTTCTGGGCC ATGCACATCA CTGAGGCTCT
 401 GCACCAGAAC ATGCAGGCTC TGTTTAGCAC CCTGGCTCAG GCGGAGGAGC
 451 AGCAGCCCTA CCTGGAGGCT CCACCGTTAT GCGCGGGACT CGCTGTCTGG
 501 CAGAGTACCA CCTGGGGGAT TATGGACACG CCTGGAACAG GTGTTGGGTG
 551 CTGGACAGGG TGGACACCTG GGCTGTGGTC ATGTTCATTG ATTTTGGACA
 601 GTTGGCCACC ATCCCTGTGC AGTCTCTGCG CCAGCTAGAC AGCGACGACT
 651 TCTGGACCAT CCCACCCCTG ACTCAGCCAT TCATGCTGGA GAAAGACATT
 701 TTGAGTTCGT ATGAGGTTGT CCATCGAATC CTCAAAGGGA AAATCACTGG
 751 TGCTTTGAAC TCGGCGGTAA CTGCTCCTGC ATCTAACTTG GCTGTTGTCC
 801 CTCCACTCCT GCCCTTGGGG TGTCTGCAGC AGGCTGCTGC CTAGGCCTGG
 851 ACACATTGCA CATCCTAAAG TTTGAAGAGT CTAAATAACG GGGCTTCCCT
 901 CAGCATGTTC CCTCTCCTGT TTGCCACGGA TCCAGAGCCA CCTGCCCTGT
 951 CTTCTCGTAC CCCTTTCACT CTTGAGGCCT GGGAGGTGAA AAAGGCCAGA
1001 CTGTGCCCAG GATTGATTCA ATTTTGCTTT TACTCCCAGC TTCCCTCTCA
1051 AAAGAGAGTG AAGTCTCATT TGTCATGTGT CTTCAGTTCC CCAACTTGGC
1101 ATGAACATTT GAACCAAACA TAGGAAACTA CCATTAGGTT GAAAGCCTGA
1151 GGCAGCTGGG ATGGTCTTTC TTGTGTCTCT TCTTTGCACC CCAGAGCATG
1201 ATATAAGTGG TCCTAACAGA TTCTGGATAA TGGAGAAGCC CTCTGCTGGT
1251 TTTCCTGGCA TTCCATGTAG AATAGGTAGA GAATATTTAA CCAATGAGCA
1301 AATAAATGTT GGCATGTTTC ATGAAAAAAA AAAAAAAAAA A
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 479 bp to 841 bp; peptide length: 121
Category: questionable ORF
Classification: no clue

```
  1 MRGTRCLAEY HLGDYGHAWN RCWVLDRVDT WAVVMFIDFG QLATIPVQSL
 51 RQLDSDDFWT IPPLTQPFML EKDILSSYEV VHRILKGKIT GALNSAVTAP
101 ASNLAVVPPL LPLGCLQQAA A
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2m20, frame 2

No Alert BLASTP hits found

Peptide information for frame 3

ORF from 87 bp to 635 bp; peptide length: 183
Category: putative protein
Classification: no clue

```
  1 MIQQPRAPLV LEKASGEGFG KTAAIIQLAP KAPVDLCETE KLRAAFFAVP
 51 LEMRGSFLVL LLRECFRDLS WLALIHSVRG EAGLLVTSIV PKTPFFWAMH
101 ITEALHQNMQ ALFSTLAQAE EQQPYLEAPP LCAGLAVWQS TTWGIMDTPG
151 TGVGCWTGWT PGLWSCSLIL DSWPPSLCSL CAS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2m20, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_2m20, frame 2

Report for DKFZphtes3_2m20.2

```
[LENGTH]    121
[MW]        13436.69
[pI]        5.81
[KW]        Alpha_Beta

SEQ   MRGTRCLAEYHLGDYGHAWNRCWVLDRVDTWAVVMFIDFGQLATIPVQSLRQLDSDDFWT
PRD   ccchhhhhcccccccccccccceeeeccccccccceeeeeeccccccccccccccccccccc

SEQ   IPPLTQPFMLEKDILSSYEVVHRILKGKITGALNSAVTAPASNLAVVPPLLPLGCLQQAA
PRD   cccccchhhhhhcchhhhhhhhhhcccccchhhhhhccccccceeeecccccccccccc

SEQ   A
PRD   c
```

(No Prosite data available for DKFZphtes3_2m20.2)

(No Pfam data available for DKFZphtes3_2m20.2)

Pedant information for DKFZphtes3_2m20, frame 3

Report for DKFZphtes3_2m20.3

```
[LENGTH]    183
[MW]        19971.49
[pI]        5.31
[KW]        Alpha_Beta
```

```
SEQ     MIQQPRAPLVLEKASGEGFGKTAAIIQLAPKAPVDLCETEKLRAAFFAVPLEMRGSFLVL
PRD     ccccccccceeeeccccccccccccccccccccccchhhhhhhhhhhhhhhhcchhhhhh

SEQ     LLRECFRDLSWLALIHSVRGEAGLLVTSIVPKTPFFWAMHITEALHQNMQALFSTLAQAE
PRD     hhhhhhcchhhhhhhhhhhcccceeeeeeeecccccchhhhhhhhhhhhhhhhhhhhhhh

SEQ     EQQPYLEAPPLCAGLAVWQSTTWGIMDTPGTGVGCWTGWTPGLWSCSLILDSWPPSLCSL
PRD     hhhcccccccccccceeeecccceeecccccccccccccccccccceeeecccccccceee

SEQ     CAS
PRD     ccc
```

(No Prosite data available for DKFZphtes3_2m20.3)

(No Pfam data available for DKFZphtes3_2m20.3)

DKFZphtes3_2n9

group: testes derived

DKFZphtes3_2n9 encodes a novel 184 amino acid protein with very weak similarity to Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

on genomic level encoded by HS1186N24, no splice pattern but EST matches

Sequenced by EMBL

Locus: unknown

Insert length: 1000 bp
Poly A stretch at pos. 988, polyadenylation signal at pos. 970

```
  1 CAACTTTTTA AAGATGTGAA TTGGACAGCC AGACTTGCTT ATTTGTCTGA
 51 TATCTTCAGT ATTTTTTAAT GATCTTAATG CTTCTATGCA AGGGAAGAAT
101 GCAACTTATT TTTCAATGGC AGATAAAGTT GAAGGACAAA AACAGAAGTT
151 AGAAGCTTGG AAAAACAGAA TTTCTACAGA TTGTTATGAC ATGTTTCATA
201 ATTTAACAAC AATTATCAAT GAAGTAGGTA ATGATCTTGA TATTGCACAT
251 CTGCGAAAAG TTATCAGTGA ACATCTTACA AATTTGTTAG AATGTTTTGA
301 ATTTTATTTT CCATCAAAAG AAGATCCACG CATAGGAAAT TTGTGGATCC
351 AAAATCCATT TCTTTCATCA AAAGATAACT TAAATTTAAC TGTAACTCTA
401 CAGGATAAGT TGTTGAAGCT GGCTACCGAC GAAGGATTGA AAATCAGTTT
451 TGAAAATACA GCATCACTTC CTTCATTTTG GATAAAAGCT AAAAATGACT
501 ATCCTGAGCT TGCTGAGATT GCTTTAAAAT TGCTGCTTCT TTTCCCCTCA
551 ACATACCTCT GTGAGACCGG ATTCTCTACT TTAAGTGTTA TTAAAACAAA
601 ACATAGAAAC AGTTTAAATA TACATTATCC CCTGAGGTAG CATTGTCATC
651 AATCCAACCT AGATTAGACA AATTAACAAG CAAGAAGCAA GCTCACTTAT
701 CACATTAAAA GCTTTAAATA TTGATATGTA AGGTATTGGT TCAAAGTATG
751 CATATAAGCA TTGAGTGTGA GGAATTTGCT ATTTCACTTT AAACTTTCTG
801 TCTAGTTACA GTTATGGAAG TATGAGAAGT TATGAGTGAA ACAGCAATTT
851 TCTATATAAA TTGCCTATAT GTATATTTTC AATTAAGAAT GTGTACAGTT
901 TTTATAATTC TATTTTTCCT CATATTTGTC GTATTTATTA AAATATAATT
951 TTAAATCTGT TGATTCTAAT ATTAAAACAT TTGATCTTAA AAAAAAAAAA
```

BLAST Results

Entry HS1186N24 from database EMBLNEW:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 1186N24
Score = 4921, P = 5.8e-215, identities = 989/992

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 86 bp to 637 bp; peptide length: 184
Category: similarity to unknown protein
Classification: no clue

```
  1 MQGKNATYFS MADKVEGQKQ KLEAWKNRIS TDCYDMFHNL TTIINEVGND
 51 LDIAHLRKVI SEHLTNLLEC FEFYFPSKED PRIGNLWIQN PFLSSKDNLN
101 LTVTLQDKLL KLATDEGLKI SFENTASLPS FWIKAKNDYP ELAEIALKLL
151 LLFPSTYLCE TGFSTLSVIK TKHRNSLNIH YPLR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_2n9, frame 2

TREMBLNEW:AC004883_3 gene: "WUGSC:H_DJ0771P04.2"; Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23, complete sequence., N = 1, Score = 94, P = 0.042

```
>TREMBLNEW:AC004883_3 gene: "WUGSC:H_DJ0771P04.2"; Homo sapiens PAC clone
     DJ0771P04 from 7q11.21-q11.23, complete sequence.
            Length = 533

  HSPs:

 Score = 94 (14.1 bits), Expect = 4.3e-02, P = 4.2e-02
 Identities = 39/177 (22%), Positives = 75/177 (42%)

Query:     1 MQGKNATYFSMADKVEGQKQKLEAWKNRISTDCYDMFHNLTTIINEVGNDLD-IAHLRKV 59
              +QG +     M D +     KL  W+  ++ +      F  L          + L+ I  + ++
Sbjct:   354 LQGHSQIVTQMYDLIRAFLAKLCLWETHLTRNNLAHFPTLKLASRNESDGLNYIPKIAEL 413

Query:    60 ISEHLTNLLECFEFYFPSKEDPRIGNLWIQNPFLSSKDNLNLTVTLQDKLLKLATDEGLK 119
              +E    L + F+ Y    + + +      +PF +  D+++    LQ +++ L  +  LK
Sbjct:   414 KTEFQKRLSD-FKLY---ESELTL----FSSPFSTKIDSVH--EELQMEVIDLQCNTVLK 463

Query:   120 ISFENTASLPSFWIKAKNDYPXXXXXXXXXXXXFPSTYLCETGFSTLSVIKTKHRNSL 177
             ++     +P F+         YP              F STY+CE  FS + + KTK+ + L
Sbjct:   464 TKYDKVG-IPEFYKYLWGSYPKYKHHCAKILSMFGSTYICEQLFSIMKLSKTKYCSQL 520
```

Pedant information for DKFZphtes3_2n9, frame 2

Report for DKFZphtes3_2n9.2

```
[LENGTH]     184
[MW]         21203.53
[pI]         6.52
[KW]         Alpha_Beta
[KW]         LOW_COMPLEXITY      6.52 %

SEQ    MQGKNATYFSMADKVEGQKQKLEAWKNRISTDCYDMFHNLTTIINEVGNDLDIAHLRKVI
SEG    ............................................................
PRD    cccccchhhhhhhhhhhhhhhhhhhhhhhhcchhhhhcccceeecccccccchhhhhhhh

SEQ    SEHLTNLLECFEFYFPSKEDPRIGNLWIQNPFLSSKDNLNLTVTLQDKLLKLATDEGLKI
SEG    ............................................................
PRD    hhhhhhhhhhhhcccccccccccceeeecccccccccccceeeeehhhhhhhhhhhcccee

SEQ    SFENTASLPSFWIKAKNDYPELAEIALKLLLLFPSTYLCETGFSTLSVIKTKHRNSLNIH
SEG    ....................xxxxxxxxxxxx............................
PRD    eecccccccceeeeecccchhhhhhhhhhhhhcccccccccccceeeeeeeccccccceeec

SEQ    YPLR
SEG    ....
PRD    cccc
```

(No Prosite data available for DKFZphtes3_2n9.2)

(No Pfam data available for DKFZphtes3_2n9.2)

DKFZphtes3_30f4
----

group: testes derived

DKFZphtes3_30f4 encodes a novel 192 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by LMU

Locus: /map="717.2-8 cR from top of Chr8 linkage group"

Insert length: 1388 bp
Poly A stretch at pos. 1330, polyadenylation signal at pos. 1310

```
   1 CACTGAGCCC TCCTCAGATG GTTAGTGGCT TCCAACAGCC ATCAGGAGTG
  51 TTTCTTGAAT GCCCCAGGTG TGGAGGACTT GGTCTGTGAC CACCTAGAAC
 101 CCCAGAGCTG AACAGGAAGC CGTCCCTGCA GCAACAAGAG GGCTGGAAGG
 151 GGGAGCTGCA GGCCACCCTC GGCTCTCCCA CTGCTGGGGC GGTGATGTTC
 201 GGGTGACATG TTTGAAAAAT ACTCTTAAAG ATACCAACTG TTCCCTTATA
 251 TGGCTAATGG TTTGTGCAGC CACCAGCGAT GGCGGCCCCT ATTAGAGACC
 301 AGGTTTGTTA AAACACCAAA TATTGCTGTC CACACTAGAC ATTAACCGGC
 351 TTCAGAAAAG ATGGACACCT TTTCCCACGC TGTTTCGCTT CTTAACTTTG
 401 GTCCAGCTTT AGCCACCACA CAGCGTGTGA GGGACTGCTG CTGCGGAGTC
 451 AGCCTCGTTT GTCCCTCCGC CTCCCACCAG CATGCGCCGC TTCTGAGAGA
 501 CACCAGCTCC CTGCCTCCAA GCCTGGTGCC ACAGGCCTGT CGTGAGGGAC
 551 CCCTGCTTCC GAGAGCTCCT GGGGGGGTTC TGCCCTTCAC CACCTGGGAG
 601 AGGTGTCAGT TCAGTTCCGA GTTGAACAAG GCCCGTGCAC ACAGCATGTT
 651 GGGGGCCCAG CCCAAAGTTC TTGTCACCTC CTCATGCAAA GCCAGCCATC
 701 ACCCTCCGGC CAGAGCTCAA GGTGGCCCCT TGGCCAGCCC CTCCTTGGGT
 751 CCTCCAGGAG GACTGAGCAC CCCTCCTAGC GGCATCCCTT GCCCTCCACA
 801 GTGCTGCCAG GGGCACGTCG CTCTGTGCCG TGGACTGAGA CCATCCCCTG
 851 GTGACAGAAT GACCCGTTTG TTGGAAATGC CTCGTTGCCA GAGAAACTCC
 901 CCAGGCATCT CGGAACGAAA CTATTTAGTT CCATTGTGAA CTGGCCACGG
 951 GACAGCTTTT TATCAACTTA TTAAGTTGGA GCACTGTAAT CGCGCTTGCT
1001 GAGTTAGCAG TGGTGGTAAG CGTGTGTTAA ACACATAATG TTACGTTTTA
1051 GGAGAGAGAG GTCGTAAGGA AGTGTCGTGT CGCTCATGAC TCTCTTCTAT
1101 TAGTTGGGTA ACAGTGGCCT CATGTTTGTG TCTGTGTGTA CACAGAGCCC
1151 TTAGGTTCTG CTCTGTTTCT TTGCCAGGTG AATGTTTGTG GCATGCGCTG
1201 CTGTCCGCGC CCCTCTGTCC TGCGCAGGGT TCAGCTGTGC GGCGCCCTGA
1251 TTTCCTCCAT GCACACAGAA CCTCCTTGTG TCTGTTTCTC TGTTCCTCTG
1301 TGGCTGACTC AATAAACTTT TCCCTCTGAC ATGAAAAAAA AAAAAAAAAA
1351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAG
```

BLAST Results

Entry HS548358 from database EMBL:
human STS EST67250.
Score = 2126, P = 1.5e-89, identities = 444/472

Entry HS670351 from database EMBL:
human STS WI-18501.
Score = 2089, P = 7.1e-88, identities = 445/476

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 361 bp to 936 bp; peptide length: 192
Category: putative protein
Classification: no clue

```
  1 MDTFSHAVSL LNFGPALATT QRVRDCCCGV SLVCPSASHQ HAPLLRDTSS
 51 LPPSLVPQAC REGPLLPRAP GGVLPFTTWE RCQFSSELNK ARAHSMLGAQ
101 PKVLVTSSCK ASHHPPARAQ GGPLASPSLG PPGGLSTPPS GIPCPPQCCQ
151 GHVALCRGLR PSPGDRMTRL LEMPRCQRNS PGISERNYLV PL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_30f4, frame 1

No Alert-BLASTP hits found

Pedant information for DKFZphtes3_30f4, frame 1

Report for DKFZphtes3_30f4.1

```
[LENGTH]     192
[MW]         20281.56
[pI]         9.21
[BLOCKS]     BL01013C Oxysterol-binding protein family proteins
[KW]         All_Alpha
[KW]         LOW_COMPLEXITY    10.94 %
```

```
SEQ   MDTFSHAVSLLNFGPALATTQRVRDCCCGVSLVCPSASHQHAPLLRDTSSLPPSLVPQAC
SEG   ............................................................
PRD   ccchhhhheeeccccccchhhhhhhhccccceeeecccccccccccccccccccccccccc

SEQ   REGPLLPRAPGGVLPFTTWERCQFSSELNKARAHSMLGAQPKVLVTSSCKASHHPPARAQ
SEG   ............................................................
PRD   cccccccccccccccccccchhhhhhhhhhhhhccccceeeeeccccccccccccccc

SEQ   GGPLASPSLGPPGGLSTPPSGIPCPPQCCQGHVALCRGLRPSPGDRMTRLLEMPRCQRNS
SEG   xxxxxxxxxxxxxxxxxxxx........................................
PRD   ccccccccccccccccccccccccccccchhhhhhhhcccccccchhhhhccccccccc

SEQ   PGISERNYLVPL
SEG   ............
PRD   cccccccccccc
```

(No Prosite data available for DKFZphtes3_30f4.1)

(No Pfam data available for DKFZphtes3_30f4.1)

DKFZphtes3_35b4
----

group: cell cycle

DKFZphtes3_35b4 encodes a novel 1780 amino acid protein which is C-terminal identical to human M-phase phosphoprotein-1 (MPP1).

The novel protein contains a N-terminal Pfam kinesin motor domain and a ATP/GTP-binding site motif A (P-loop). MPP1 is expressed and phosphorylated in the metaphase. Therefore the novel protein seems to be involved in the mitotic spindle during cell division.

The new protein can find application in modulation of the mitotic spindle.

"M-phase phosphoprotein-1" extension

motor protein

Sequenced by DKFZ

Locus: /map="750_H_1; 758_H_7; 759_C_9; 847_D_4; 906_D_1; 931_D_3; 944_C_1; 750_G_12; 800_A_11; 512.1 cR from top of Chr10 linkage group"

Insert length: 6284 bp
No poly A stretch found, no polyadenylation signal found

```
   1 ATCGCAGTGC TGCTCGCGGG TCTGGCTAGT CAGGCGAAGT TTGCAGAATG
  51 GAATCTAATT TTAATCAAGA GGGAGTACCT CGACCATCTT ATGTTTTTAG
 101 TGCTGACCCA ATTGCAAGGC CTTCAGAAAT AAATTTCGAT GGCATTAAGC
 151 TTGATCTGTC TCATGAATTT TCCTTAGTTG CTCCAAATAC TGAGGCAAAC
 201 AGTTTCGAAT CTAAAGATTA TCTCCAGGTT TGTCTTCGAA TAAGACCATT
 251 TACACAGTCA GAAAAAGAAC TTGAGTCTGA GGGCTGTGTG CATATTCTGG
 301 ATTCACAGAC TGTTGTGCTG AAAGAGCCTC AATGCATCCT TGGTCGGTTA
 351 AGTGAAAAAA GCTCAGGGCA GATGGCACAG AAATTCAGTT TTTCCAAGGT
 401 TTTTGGCCCA GCAACTACAC AGAAGGAATT CTTTCAGGGT TGCATTATGC
 451 AACCAGTAAA AGACCTCTTG AAAGGACAGA GTCGTCTGAT TTTTACTTAC
 501 GGGCTAACCA ATTCAGGAAA AACATATACA TTTCAAGGGA CAGAAGAAAA
 551 TATTGGCATT CTGCCTCGAA CTTTGAATGT ATTATTTGAT AGTCTTCAAG
 601 AAAGACTGTA TACAAAGATG AACCTTAAAC CACATAGATC CAGAGAATAC
 651 TTAAGGTTAT CATCAGAACA AGAGAAAGAA GAAATTGCTA GCAAAAGTGC
 701 ATTGCTTCGG CAAATTAAAG AGGTTACTGT GCATAATGAT AGTGATGATA
 751 CTCTTTATGG AAGTTTAACT AACTCTTTGA ATATCTCAGA GTTTGAAGAA
 801 TCCATAAAAG ATTATGAACA AGCCAACTTG AATATGGCTA ATAGTATAAA
 851 ATTTTCTGTG TGGGTTTCTT TCTTTGAAAT TTACAATGAA TATATTTATG
 901 ACTTATTTGT TCCTGTATCA TCTAAATTCC AAAAGAGAAA GATGCTGCGC
 951 CTTTCCCAAG ACGTAAAGGG CTATTCTTTT ATAAAAGATC TACAATGGAT
1001 TCAAGTATCT GATTCCAAAG AAGCCTATAG ACTTTTAAAA CTAGGAATAA
1051 AGCACCAGAG TGTTGCCTTC ACAAAATTGA ATAATGCTTC CAGTAGAAGT
1101 CACAGCATAT TCACTGTTAA AATATTACAG ATTGAAGATT CTGAAATGTC
1151 TCGTGTAATT CGAGTCAGTG AATTATCTTT ATGTGATCTT GCTGGTTCAG
1201 AACGAACTAT GAAGACACAG AATGAAGGTG AAAGGTTAAG AGAGACTGGG
1251 AATATCAACA CTTCTTTATT GACTCTGGGA AAGTGTATTA ACGTCTTGAA
1301 GAATAGTGAA AAGTCAAAGT TTCAACAGCA TGTGCCTTTC CGGGAAAGTA
1351 AACTGACTCA CTATTTTCAA AGTTTTTTTA ATGGTAAAGG GAAAATTTGT
1401 ATGATTGTCA ATATCAGCCA ATGTTATTTA GCCTATGATG AAACACTCAA
1451 TGTATTGAAG TTCTCCGCCA TTGCACAAAA AGTTTGTGTC CCAGACACTT
1501 TAAATTCCTC TCAAGATAAA TTATTTGGAC CTGTCAAATC TTCTCAAGAT
1551 GTATCACTAG ACAGTAATTC AAACAGTAAA ATATTAAATG TAAAAAGAGC
1601 CACCATTTCA TGGGAAAATA GTCTAGAAGA TTTGATGGAA GACGAGGATT
1651 TGGTTGAGGA GCTAGAAAAC GCTGAAGAAA CTCAAAATGT GGAAACTAAA
1701 CTTCTTGATG AAGATCTAGA TAAAACATTA GAGGAAAATA AGGCTTTCAT
1751 TAGCCACGAG GAGAAAAGAA AACTGTTGGA CTTAATAGAA GACTTGAAAA
1801 AAAAACTGAT AAATGAAAAA AAGGAAAAAT TAACCTTGGA ATTTAAAATT
1851 CGAGAAGAAG TTACACAGGA GTTTACTCAG TATTGGGCTC AACGGGAAGC
1901 TGACTTTAAG GAGACTCTGC TTCAAGAACG AGAGATATTA GAAGAAAATG
1951 CTGAACGTCG TTTGGCTATC TTCAAGGATT TGGTTGGTAA ATGTGACACT
2001 CGAGAAGAAG CAGCGAAAGA CATTTGTGCC ACAAAAGTTG AAACTGAAGA
2051 AGCTACTGCT TGTTTAGAAC TAAAGTTTAA TCAAATTAAA GCTGAATTAG
2101 CTAAAACCAA AGGAGAATTA ATCAAAACCA AAGAAGAGTT AAAAAAGAGA
2151 GAAAATGAAT CAGATTCATT GATTCAAGAG CTTGAGACAT CTAATAAGAA
2201 AATAATTACA CAGAATCAAA GAATTAAAGA ATTGATAAAT ATAATTGATC
2251 AAAAAGAAGA TACTATCAAC GAATTTCAGA ACCTAAAGTC TCATATGGAA
2301 AACACATTTA AATGCAATGA CAAGGCTGAT ACATCTTCTT TAATAATAAA
2351 CAATAAATTG ATTTGTAATG AAACAGTTGA AGTACCTAAG GACAGCAAAT
2401 CTAAAATCTG TTCAGAAAGA AAAAGAGTAA ATGAAAATGA ACTTCAGCAA
2451 GATGAACCAC CAGCAAAGAA AGGGTCTATC CATGTTAGTT CAGCTATCAC
2501 TGAAGACCAA AAGAAAAGTG AAGAAGTGCG ACCGAACATT GCAGAAATTG
2551 AAGACATCAG AGTTTTACAA GAAAATAATG AAGGACTGAG AGCATTTTTA
```

```
2601 CTCACTATTG AGAATGAACT TAAAAATGAA AAGGAAGAAA AAGCAGAATT
2651 AAATAAACAG ATTGTTCATT TTCAGCAGGA ACTTTCTCTT TCTGAAAAAA
2701 AGAATTTAAC TTTAAGTAAA GAGGTCCAAC AAATTCAGTC AAATTATGAT
2751 ATTGCAATTG CTGAATTACA TGTGCAGAAA AGTAAAAATC AAGAACAGGA
2801 GGAAAAGATC ATGAAATTGT CAAATGAGAT AGAAACTGCT ACAAGAAGCA
2851 TTACAAATAA TGTTTCACAA ATAAAATTAA TGCACACGAA AATAGACGAA
2901 CTACGTACTC TTGATTCAGT TTCTCAGATT TCAAACATAG ATTTGCTCAA
2951 TCTCAGGGAT CTGTCAAATG GTTCTGAGGA GGATAATTTG CCAAATACAC
3001 AGTTAGACCT TTTAGGTAAT GATTATTTGG TAAGTAAGCA AGTTAAAGAA
3051 TATCGAATTC AAGAACCCAA TAGGGAAAAT TCTTTCCACT CTAGTATTGA
3101 AGCTATTTGG GAAGAATGTA AAGAGATTGT GAAGGCCTCT TCCAAAAAAA
3151 GTCATCAGAT TGAGGAACTG GAACAACAAA TTGAAAAATT GCAGGCAGAA
3201 GTAAAAGGCT ATAAGGATGA AAACAATAGA CTAAAGGAGA AGGAGCATAA
3251 AAACCAAGAT GACCTACTAA AAGAAAAAGA AACTCTTATA CAGCAGCTGA
3301 AAGAAGAATT GCAAGAAAAA AATGTTACTC TTGATGTTCA AATACAGCAT
3351 GTAGTTGAAG GAAAGAGAGC GCTTTCAGAA CTTACACAAG GTGTTACTTG
3401 CTATAAGGCA AAAATAAAGG AACTTGAAAC AATTTTAGAG ACTCAGAAAG
3451 TTGAACGTAG TCATTCAGCC AAGTTAGAAC AAGACATTTT GGAAAAGGAA
3501 TCTATCATCT TAAAGCTAGA AAGAAATTTG AAGGAATTTC AAGAACATCT
3551 TCAGGATTCT GTCAAAAACA CCAAAGATTT AAATGTAAAG GAACTCAAGC
3601 TGAAAGAAGA AATCACACAG TTAACAAATA ATTTGCAAGA TATGAAACAT
3651 TTACTTCAAT TAAAAGAAGA AGAAGAAGAA ACCAACAGGC AAGAAACAGA
3701 AAAATTGAAA GAGGAACTCT CTGCAAGCTC TGCTCGTACC CAGAATCTGA
3751 AAGCAGATCT TCAGAGGAAG GAAGAAGATT ATGCTGACCT GAAAGAGAAA
3801 CTGACTGATG CCAAAAAGCA GATTAAGCAA GTACAGAAAG AGGTATCTGT
3851 AATGCGTGAT GAGGATAAAT TACTGAGGAT TAAAATTAAT GAACTGGAGA
3901 AAAAGAAAAA CCAGTGTTCT CAGGAATTAG ATATGAAGCA GCGAACCATT
3951 CAGCAACTCA AGGAGCAGTT AAATAATCAG AAAGTGGAAG AAGCTATACA
4001 ACAGTATGAG AGAGCATGCA AAGATCTAAA TGTTAAAGAG AAAATAATTG
4051 AAGACATGCG AATGACACTA GAAGAACAGG AACAAACTCA GGTAGAACAG
4101 GATCAAGTGC TTGAGGCTAA ATTAGAGGAA GTTGAAAGGC TGGCCACAGA
4151 ATTGGAAAAA TGGAAGGAAA AATGCAATGA TTTGGAAACC AAAAACAATC
4201 AAAGGTCAAA TAAAGAACAT GAGAACAACA CAGATGTGCT TGGAAAGCTC
4251 ACTAATCTTC AAGATGAGTT ACAGGAGTCT GAACAGAAAT ATAATGCTGA
4301 TAGAAAGAAA TGGTTAGAAG AAAAAATGAT GCTTATCACT CAAGCGAAAG
4351 AAGCAGAGAA TATACGAAAT AAAGAGATGA AAAAATATGC TGAGGACAGG
4401 GAGCGTTTTT TTAAGCAACA GAATGAAATG GAAATACTGA CAGCCCAGCT
4451 GACAGAGAAA GATAGTGACC TTCAAAAGTG GCGAGAAGAA CGAGATCAAC
4501 TGGTTGCAGC TTTAGAAATA CAGCTAAAAG CACTGATATC CAGTAATGTA
4551 CAGAAAGATA ATGAAATTGA ACAACTAAAA AGGATCATAT CAGAGACTTC
4601 TAAAATAGAA ACACAAATCA TGGATATCAA GCCCAAACGT ATTAGTTCAG
4651 CAGATCCTGA CAAACTTCAA ACTGAACCTC TATCGACAAG TTTTGAAATT
4701 TCCAGAAATA AAATAGAGGA TGGATCTGTA GTCCTTGACT CTTGTGAAGT
4751 GTCAACAGAA AATGATCAAA GCACTCGATT TCCAAAACCT GAGTTAGAGA
4801 TTCAATTTAC ACCTTTACAG CCAAACAAAA TGGCAGTGAA ACACCCTGGT
4851 TGTACCACAC CAGTGACAGT TGAGATTCCC AAGGCTCGGA AGAGGAAGAG
4901 TAATGAAATG GAGGAGGACT TGGTGAAATG TGAAAATAAG AAGAATGCTA
4951 CACCCAGAAC TAATTTGAAA TTTCCTATTT CAGATGATAG AAATTCTTCT
5001 GTCAAAAAGG AACAAAAGGT TGCCATACGT CCATCATCTA AGAAAACATA
5051 TTCTTTACGG AGTCAGGCAT CCATAATTGG TGTAAACCTG GCCACTAAGA
5101 AAAAAGAAGG AACACTACAG AAATTTGGAG ACTTCTTACA ACATTCTCCC
5151 TCAATTCTTC AATCAAAAGC AAAGAAGATA ATTGAAACAA TGAGCTCTTC
5201 AAAGCTCTCA AATGTAGAAG CAAGTAAAGA AAATGTGTCT CAACCAAAAC
5251 GAGCCAAACG GAAATTATAC ACAAGTGAAA TTTCATCTCC TATTGATATA
5301 TCAGGCCAAG TGATTTTAAT GGACCAGAAA ATGAAGGAGA GTGATCACCA
5351 GATTATCAAA CGACGACTTC GAACAAAAAC AGCCAAATAA ATCACTTATG
5401 GAAATGTTTA ATATAAATTT TATAGTCATA GTCATTGGAA CTTGCATCCT
5451 GTATTGTAAA TATAAATGTA TATATTATGC ATTAAATCAC TCTGCATATA
5501 GATTGCTGTT TTATACATAG TATAATTTTA ATTCAATAAA TGAGTCAAAA
5551 TTTGTATATT TTTATAAGGC TTTTTATAAA TAGCTTCTTT CAAACTGTAT
5601 TTCCCTATTA TCTCAGACAT TGGATCAGTG AAGATCCTAG GAAAGAGGCT
5651 GTTATTCTCA TTTATTTTGC TATACAGGAT GTAATAGGTC AGGTATTTGG
5701 TTTACTTATA TTTAACAATG TCTTATGAAT TTTTTTTACT TTATCTGTTA
5751 TACAACTGAT TTTACATATC TGTTTGGATT ATAGCTAGGA TTTGGAGAAT
5801 AAGTGTGTAC AGATCACAAA ACATGTATAT ACATTATTTA GAAAAGATCT
5851 CAAGTCTTTA ATTAGAATGT CTCACTTATT TTGTAAACAT TTTGTGGGTA
5901 CATAGTACAT GTATATATTT ACGGGGTATG TGAGATGTTT TGACACAGGC
5951 ATGCAATGTG AAATACGTGT ATCATGGAGA ATGAGGTATC CATCCCCTCA
6001 AGCATTTTTC CTTTGAATTA CAGATAATCC AATTACATTC TTTAGATCAT
6051 TTAAAAATAT ACAAGTAAGT TATTATTGAT TATAGTCACT CTATTGTGCT
6101 ATCAGATAGT AGATCATTCT TTTTATCTTA TTTGTTTTTG TACCCATTAA
6151 CCATCCCCAC CTCCCCCTGC AACCGTCAGT ACCCTTACCA GCCACTGGTA
6201 ACCATTCTTC TACTCTGTAT GCCCATGAGG TCAATTGATT TTATTTTTAG
6251 ATCCCATAAA TAAATGAGAA CATGCAAAAA AAAA
```

BLAST Results

Entry HS898149 from database EMBL:
human STS WI-9217.

Score = 4247, P = 1.5e-187, identities = 855/862

Medline entries

94119956:
Cloning of cDNAs for M-phase phosphoproteins recognized by the MPM2 monoclonal antibody and determination of the phosphorylated epitope.

98101856:
Interaction of a Golgi-associated kinesin-like protein with Rab6.

95122643:
Identification and partial characterization of mitotic centromere-associated kinesin, a
kinesin-related protein that associates with centromeres during mitosis.

Peptide information for frame 3

ORF from 48 bp to 5387 bp; peptide length: 1780
Category: known protein
Classification: Cell structure/motility
Prosite motifs: ATP_GTP_A (152-160)

```
   1 MESNFNQEGV PRPSYVFSAD PIARPSEINF DGIKLDLSHE FSLVAPNTEA
  51 NSFESKDYLQ VCLRIRPFTQ SEKELESEGC VHILDSQTVV LKEPQCILGR
 101 LSEKSSGQMA QKFSFSKVFG PATTQKEFFQ GCIMQPVKDL LKGQSRLIFT
 151 YGLTNSGKTY TFQGTEENIG ILPRTLNVLF DSLQERLYTK MNLKPHRSRE
 201 YLRLSSEQEK EEIASKSALL RQIKEVTVHN DSDDTLYGSL TNSLNISEFE
 251 ESIKDYEQAN LNMANSIKFS VWVSFFEIYN EYIYDLFVPV SSKFQKRKML
 301 RLSQDVKGYS FIKDLQWIQV SDSKEAYRLL KLGIKHQSVA FTKLNNASSR
 351 SHSIFTVKIL QIEDSEMSRV IRVSELSLCD LAGSERTMKT QNEGERLRET
 401 GNINTSLLTL GKCINVLKNS EKSKFQQHVP FRESKLTHYF QSFFNGKGKI
 451 CMIVNISQCY LAYDETLNVL KFSAIAQKVC VPDTLNSSQD KLFGPVKSSQ
 501 DVSLDSNSNS KILNVKRATI SWENSLEDLM EDEDLVEELE NAEETQNVET
 551 KLLDEDLDKT LEENKAFISH EEKRKLLDLI EDLKKKLINE KKEKLTLEFK
 601 IREEVTQEFT QYWAQREADF KETLLQEREI LEENAERRLA IFKDLVGKCD
 651 TREEAAKDIC ATKVETEEAT ACLELKFNQI KAELAKTKGE LIKTKEELKK
 701 RENESDSLIQ ELETSNKKII TQNQRIKELI NIIDQKEDTI NEFQNLKSHM
 751 ENTFKCNDKA DTSSLIINNK LICNETVEVP KDSKSKICSE RKRVNENELQ
 801 QDEPPAKKGS IHVSSAITED QKKSEEVRPN IAEIEDIRVL QENNEGLRAF
 851 LLTIENELKN EKEEKAELNK QIVHFQQELS LSEKKNLTLS KEVQQIQSNY
 901 DIAIAELHVQ KSKNQEQEEK IMKLSNEIET ATRSITNNVS QIKLMHTKID
 951 ELRTLDSVSQ ISNIDLLNLR DLSNGSEEDN LPNTQLDLLG NDYLVSKQVK
1001 EYRIQEPNRE NSFHSSIEAI WEECKEIVKA SSKKSHQIEE LEQQIEKLQA
1051 EVKGYKDENN RLKEKEHKNQ DDLLKEKETL IQQLKEELQE KNVTLDVQIQ
1101 HVVEGKRALS ELTQGVTCYK AKIKELETIL ETQKVERSHS AKLEQDILEK
1151 ESIILKLERN LKEFQEHLQD SVKNTKDLNV KELKLKEEIT QLTNNLQDMK
1201 HLLQLKEEEE ETNRQETEKL KEELSASSAR TQNLKADLQR KEEDYADLKE
1251 KLTDAKKQIK QVQKEVSVMR DEDKLLRIKI NELEKKKNQC SQELDMKQRT
1301 IQQLKEQLNN QKVEEAIQQY ERACKDLNVK EKIIEDMRMT LEEQEQTQVE
1351 QDQVLEAKLE EVERLATELE KWKEKCNDLE TKNNQRSNKE HENNTDVLGK
1401 LTNLQDELQE SEQKYNADRK KWLEEKMMLI TQAKEAENIR NKEMKKYAED
1451 RERFFKQQNE MEILTAQLTE KDSDLQKWRE ERDQLVAALE IQLKALISSN
1501 VQKDNEIEQL KRIISETSKI ETQIMDIKPK RISSADPDKL QTEPLSTSFE
1551 ISRNKIEDGS VVLDSCEVST ENDQSTRFPK PELEIQFTPL QPNKMAVKHP
1601 GCTTPVTVEI PKARKRKSNE MEEDLVKCEN KKNATPRTNL KFPISDDRNS
1651 SVKKEQKVAI RPSSKKTYSL RSQASIIGVN LATKKKEGTL QKFGDFLQHS
1701 PSILQSKAKK IIETMSSSKL SNVEASKENV SQPKRAKRKL YTSEISSPID
1751 ISGQVILMDQ KMKESDHQII KRRLRTKTAK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35b4, frame 3

TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase phosphoprotein-1 mRNA, partial cds., N = 1, Score = 3743, P = 0

```
PIR:A36881 MPM2-reactive phosphoprotein 1 - human (fragment), N = 2,
Score = 2808, P = 2.5e-294

TREMBL:AF070672_1 product: "rabkinesin6"; Homo sapiens rabkinesin6
mRNA, complete cds., N = 2, Score = 680, P = 2.6e-99


>TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase
     phosphoprotein-1 mRNA, partial cds.
            Length = 753

  HSPs:

 Score = 3743 (561.6 bits), Expect = 0.0e+00, P = 0.0e+00
 Identities = 752/753 (99%), Positives = 753/753 (100%)

Query:  1028 VKASSKKSHQIEELEQQIEKLQAEVKGYKDENNRLKEKEHKNQDDLLKEKETLIQQLKEE 1087
             VKASSKKSHQIEELEQQIEKLQAEVKGYKDENNRLKEKEHKNQDDLLKEKETLIQQLKEE
Sbjct:     1 VKASSKKSHQIEELEQQIEKLQAEVKGYKDENNRLKEKEHKNQDDLLKEKETLIQQLKEE 60

Query:  1088 LQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 1147
             LQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI
Sbjct:    61 LQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 120

Query:  1148 LEKESIILKLERNLKEFQEHLQDSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLLQLKE 1207
             LEKESIILKLERNLKEFQEHLQDSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLLQLKE
Sbjct:   121 LEKESIILKLERNLKEFQEHLQDSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLLQLKE 180

Query:  1208 EEEETNRQETEKLKEELSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS 1267
             EEEETNRQETEKLKEELSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS
Sbjct:   181 EEEETNRQETEKLKEELSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS 240

Query:  1268 VMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQQYERACKDL 1327
             VMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQQYERACKDL
Sbjct:   241 VMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQQYERACKDL 300

Query:  1328 NVKEKIIEDMRMTLEEQEQTQVEQDQVLEAKLEEVERLATELEKWKEKCNDLETKNNQRS 1387
             NVKEKIIEDMRMTLEEQEQTQVEQDQVLEAKLEEVERLATELEKWKEKCNDLETKNNQRS
Sbjct:   301 NVKEKIIEDMRMTLEEQEQTQVEQDQVLEAKLEEVERLATELEKWKEKCNDLETKNNQRS 360

Query:  1388 NKEHENNTDVLGKLTNLQDELQESEQKYNADRKKWLEEKMMLITQAKEAENIRNKEMKKY 1447
             NKEHENNTDVLGKLTNLQDELQESEQKYNADRKKWLEEKMMLITQAKEAENIRNKEMKKY
Sbjct:   361 NKEHENNTDVLGKLTNLQDELQESEQKYNADRKKWLEEKMMLITQAKEAENIRNKEMKKY 420

Query:  1448 AEDRERFFKQQNEMEILTAQLTEKDSDLQKWREERDQLVAALEIQLKALISSNVQKDNEI 1507
             AEDRERFFKQQNEMEILTAQLTEKDSDLQKWREERDQLVAALEIQLKALISSNVQKDNEI
Sbjct:   421 AEDRERFFKQQNEMEILTAQLTEKDSDLQKWREERDQLVAALEIQLKALISSNVQKDNEI 480

Query:  1508 EQLKRIISETSKIETQIMDIKPKRISSADPDKLQTEPLSTSFEISRNKIEDGSVVLDSCE 1567
             EQLKRIISETSKIETQIMDIKPKRISSADPDKLQTEPLSTSFEISRNKIEDGSVVLDSCE
Sbjct:   481 EQLKRIISETSKIETQIMDIKPKRISSADPDKLQTEPLSTSFEISRNKIEDGSVVLDSCE 540

Query:  1568 VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTVEIPKARKRKSNEMEEDLVK 1627
             VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTV+IPKARKRKSNEMEEDLVK
Sbjct:   541 VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTVKIPKARKRKSNEMEEDLVK 600

Query:  1628 CENKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPSSKKTYSLRSQASIIGVNLATKKKE 1687
             CENKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPSSKKTYSLRSQASIIGVNLATKKKE
Sbjct:   601 CENKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPSSKKTYSLRSQASIIGVNLATKKKE 660

Query:  1688 GTLQKFGDFLQHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS 1747
             GTLQKFGDFLQHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS
Sbjct:   661 GTLQKFGDFLQHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS 720

Query:  1748 PIDISGQVILMDQKMKESDHQIIKRRLRTKTAK 1780
             PIDISGQVILMDQKMKESDHQIIKRRLRTKTAK
Sbjct:   721 PIDISGQVILMDQKMKESDHQIIKRRLRTKTAK 753

 Score = 197 (29.6 bits), Expect = 2.1e-11, P = 2.1e-11
 Identities = 114/542 (21%), Positives = 253/542 (46%)

Query:   692 IKTKEELKKRENESDSLIQELETSNKKIITQNQRIKELINIIDQKEDTINEFQNLKSHM- 750
             +K   +   + E +  I++L+   K    +N R+KE  +    ++D + E + L   +
Sbjct:     1 VKASSKKSHQIEELEQQIEKLQAEVKGYKDENNRLKEKEH--KNQDDLLKEKETLIQQLK 58

Query:   751 ENTFKCNDKADTS-SLIINNKLICNETVEVPKDSKSKICSERKRVNENELQQDEPPAK-- 807
             E   + N   D    ++  K   +E  +      K+KI  E + + E +  +     AK
Sbjct:    59 EELQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKI-KELETILETQKVERSHSAKLE 117

Query:   808 KGSIHVSSAITEDQKKSEEVRPNIAE-IEDIRVLQENNEGLRAFLLTIENELKNEK---- 862
```

```
                 +  +    S I + ++  +E + ++ + +++ + L       L+  +   + N L++ K
Sbjct:      118 QDILEKESIILKLERNLKEFQEHLQDSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLLQ 177

Query:      863 --EEKAELNKQIVH-FQQELSLSEKKNLTLSKEVQQIQSNYDIAIAELHVQKSKNQEQEE 919
                  EE+ E N+Q      ++ELS S  +    L  ++Q+ + +Y     A+L   K K  + ++
Sbjct:      178 LKEEEEETNRQETEKLKEELSASSARTQNLKADLQRKEEDY----ADL---KEKLTDAKK 230

Query:      920 KIMKLSNEIETATRSITNNVSQIKLMHTKIDEL-RTLDSVSQISNIDLLNLRDLSNGSEE 978
                +I ++  E+      S+  +  + KL+  KI+EL + +  SQ   +D+    R +     E+
Sbjct:      231 QIKQVQKEV-----SVMRD--EDKLLRIKINELEKKKNQCSQ--ELDMKQ-RTIQQLKEQ 280

Query:      979 DNLPNTQLDLLGNDYLVSKQVKEYRIQEPNRENSFHSSIEAIWEECKEIVKASSKKSHQI 1038
                 N  N +++     Y   +  K+  ++E   E+     ++E    +   E  +     K  ++
Sbjct:      281 LN--NQKVEEAIQQY--ERACKDLNVKEKIIED-MRMTLEEQEQTQVEQDQVLEAKLEEV 335

Query:     1039 EELEQQIEKLQAEVKGYKDENNRLKEKEHKNQDDLLKEKETLIQQLKEELQEKNVT---- 1094
                E L   ++EK + +      + +NN+    KEH+N  D+L +    L  +L+E  Q+ N
Sbjct:      336 ERLATELEKWKEKCNDLETKNNQRSNKEHENNTDVLGKLTNLQDELQESEQKYNADRKKW 395

Query:     1095 LDVQIQHVVEGKRA-------LSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 1147
                L+ ++  + + K A         + + +        +  +  E+E IL  Q  E+      + ++
Sbjct:      396 LEEKMMLITQAKEAENIRNKEMKKYAEDRERFFKQQNEME-ILTAQLTEKDSDLQKWRE- 453

Query:     1148 LEKESIILKLERNLKEFQEHLQDSVKNTKDLNVKELK-LKEEITQLTNNLQDMKHLLQLK 1206
                 E++ ++  LE  LK      +  +V+  KD  +++LK +  E +++    + D+K       +
Sbjct:      454 -ERDQLVAALEIQLKAL---ISSNVQ--KDNEIEQLKRIISETSKIETQIMDIK---PKR 504

Query:     1207 EEEEETNRQETEKLKEELSASSARTQN 1233
                   + ++ +TE L       S  + ++
Sbjct:      505 ISSADPDKLQTEPLSTSFEISRNKIED 531

 Score = 186 (27.9 bits), Expect = 3.2e-10, P = 3.2e-10
 Identities = 131/674 (19%), Positives = 294/674 (43%)

Query:      673 LELKFNQIKAELAKTKGELIKT-KEELKKRENESDSLIQELETSNKKIITQNQRIKELIN 731
                L+ K ++ + +L K K  LI+  KEEL+++    D  IQ +    + +     Q +
Sbjct:       35 LKEKEHKNQDDLLKEKETLIQQLKEELQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKA 94

Query:      732 IIDQKEDTINEFQNL-KSHMENTFKCNDKADTSSLIINNKLICNETVEVPKDSKSKICSE 790
                 I + E TI E Q + +SH        +  D  +  S+I+ +   +    E  E  +DS
Sbjct:       95 KIKELE-TILETQKVERSHSAKLEQ--DILEKESIILKLERNLKEFQEHLQDS-----VKN 147

Query:      791 RKRVNENELQ-QDEPPAKKGSIHVSSAITEDQKKSEEV-RPNIAEI-EDIRVLQENNEGL 847
                 K +N  EL+ ++E       ++     + + +++ EE   R    ++ E++     + L
Sbjct:      148 TKDLNVKELKLKEEITQLTNNLQDMKHLLQLKEEEEETNRQETEKLKEELSASSARTQNL 207

Query:      848 RAFLLTIENELKNEKEEKAELNKQIVHFQQELSLSEKKNLTLSKEVQQI-----QSNYDI 902
                +A L    E +  + KE+  +  KQI   Q+E+S+   ++  L  ++  ++          Q + ++
Sbjct:      208 KADLQRKEEDYADLKEKLTDAKKQIKQVQKEVSVMRDEDKLLRIKINELEKKKNQCSQEL 267

Query:      903 AIAELHVQKSKNQEQEEKIMKLSNEIETATRSITNNVSQIKLMHTKIDEL-RTLDSVSQI 961
                 +  +  +Q+ K Q   +K+ +  +    + E A + +        I+ M   ++ E +T     Q+
Sbjct:      268 DMKQRTIQQLKEQLNNQKVEEAIQQYERACKDLNVKEKIIEDMRMTLEEQEQTQVEQDQV 327

Query:      962 SNIDLLNLRDLSNGSEEDNLPNTQLDLLGNDYLVSKQVKEYRI--QEPNRENSFHSSIEA 1019
                   L   +  L+   E+       L+   N         +         +  N ++      S +
Sbjct:      328 LEAKLEEVERLATELEKWKEKCNDLETKNNQRSNKEHENNTDVLGKLTNLQDELQESEQK 387

Query:     1020 IWEECKEIVKASSKKSHQIEELEQQIEKLQAEVKGYKDENNRLKEKEHKNQ--DDLLKEK 1077
                   + K+ ++          Q +E E    K     E+K Y ++    R  +++++ +      L EK
Sbjct:      388 YNADRKKWLEEKMMLITQAKEAENIRNK---EMKKYAEDRERFFKQQNEMEILTAQLTEK 444

Query:     1078 ETLIQQLKEELQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVER 1137
                ++ +Q+ +EE   +    L++Q++ ++             + +    ++    ++ET +    K +R
Sbjct:      445 DSDLQKWREERDQLVAALEIQLKALISSNVQKDNEIEQLKRIISETSKIETQIMDIKPKR 504

Query:     1138 SHSAKLEQDILEKESIILKLERNLKEFQEHLQDS----VKNTKDLNVKELKLKEEITQLT 1193
                  SA  ++    E  S   ++ RN  E    + DS      +N  +      +  + +L+ + + T L
Sbjct:      505 ISSADPDKLQTEPLSTSFEISRNKIEDGSVVLDSCEVSTENDQSTRFPKPELEIQFTPLQ 564

Query:     1194 NNLQDMKH---LLQLKEEEEETNRQETEKLKEEL-SASSARTQNLKADLQRKEEDYADLK 1249
                N    +KH         +   +   ++++++ +++E+L     +       + +L+     D   +
Sbjct:      565 PNKMAVKHPGCTTPVTVKIPKARKRKSNEMEEDLVKCENKKNATPRTNLKFPISDDRNSS 624

Query:     1250 EKLTDAKKQIKQVQKEVSVMRDEDKLLRIKINELEKKKNQCSQEL-DMKQRTIQQLKEQL 1308
                 K   + K   I+    K+    +R  +  +   I +N    KKK     Q+  D  Q  +    L+ +
Sbjct:      625 VK-KEQKVAIRPSSKKTYSLRSQASI--IGVNLATKKKEGTLQKFGDFLQHSPSILQSKA 681

Query:     1309 NNQKVEEAIQQYERACKDLNVKEKIIEDMR 1338
                 +K+ E +    + +  + + KE +  +  R
Sbjct:      682 --KKIIETMSSSKLSNVEAS-KENVSQPKR 708
```

```
 Score = 165 (24.8 bits), Expect = 5.8e-08, P = 5.8e-08
 Identities = 140/626 (22%), Positives = 271/626 (43%)

Query:   536 VEELENAEETQNVETKLLDEDLDKTLEENKAFISHEEKRKLLDLIEDLKKKLINEKKEK- 594
             +EELE   E    E K   +D +  L+E +    H+ +  LL    E L ++L  E +EK
Sbjct:    11 IEELEQQIEKLQAEVKGY-KDENNRLKEKE----HKNQDDLLKEKETLIQQLKEELQEKN 65

Query:   595 LTLEFKIREEVT-----QEFTQYWAQREADFKE--TLLQEREILEENAERRLAIFKDLVG 647
             +TL+ +I+  V        E TQ      +A  KE  T+L+ +++ E +    +L   +D++
Sbjct:    66 VTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKV-ERSHSAKLE--QDILE 122

Query:   648 KCDT---REEAAKDICATKVETEEATACLELKFNQIKAELAKTKGELIKTKEELKKRENE 704
             K       E  K+       ++ + T  L +K  ++K E+ +     L  K  L+ +E E
Sbjct:   123 KESIILKLERNLKEFQEHLQDSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLLQLKEEE 182

Query:   705 SDSLIQELETSNKKIITQNQRIKELINIIDQKEDTINEFQNLKSHMENTFKCNDKADTSS 764
              ++  QE E    +++    R + L    + +KE+    +   +    +    K    K + S
Sbjct:   183 ETNRQETEKLKEELSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQK-EVSV 241

Query:   765 LIINNKLICNETVEVPKDSKSKICSERKRVNENELQQDEPPAKKGSIHVSSAITEDQKKS 824
             +   +KL+  +  E+ K K  CS+   + +  +QQ +          V  AI + ++
Sbjct:   242 MRDEDKLLRIKINELEK--KKNQCSQELDMKQRTIQQLKEQLNNQK--VEEAIQQYERAC 297

Query:   825 EEVRPNIAEIEDIRVLQENNEGLRAFLLTIENELKNEKEEKAELNKQIVHFQQELSLSEK 884
             +++      IED+R+  E  E  +   +   +  L+ + EE   L ++  ++++ + E
Sbjct:   298 KDLNVKEKIIEDMRMTLEEQEQTQ---VEQDQVLEAKLEEVERLATELEKWKEKCNDLET 354

Query:   885 KNLTLSKEVQQIQSNYDIAIAELHVQKSKNQEQEEKIMKLSNE-IETATRSITN-----N 938
             KN    S +  + ++N D+ + +L    + + QE E+K       + +E    IT     N
Sbjct:   355 KNNQRSNK--EHENNTDV-LGKLTNLQDELQESEQKYNADRKKWLEEKMMLITQAKEAEN 411

Query:   939 VSQIKLMHTKIDELRTLDSVSQISNIDL-LNLRD--LSNGSEEDNLPNTQLDLLGNDYLV 995
             +   ++     D  R     +++  +   L  +D  L     EE +      L++    +
Sbjct:   412 IRNKEMKKYAEDRERFFKQQNEMEILTAQLTEKDSDLQKWREERDQLVAALEIQLKALIS 471

Query:   996 SKQVKEYRIQEPNRENSFHSSIEA-IWE-ECKEIVKASSKKSHQIEELEQQIEKLQAEVK 1053
             S   K+   I++   R   S   S IE   I  +  +K I   A   K   Q E L   E  +  +++
Sbjct:   472 SNVQKDNEIEQLKRIISETSKIETQIMDIKPKRISSADPDKL-QTEPLSTSFEISRNKIE 530

Query:  1054 GYKDENNRLKEKEHKNQDDLLKEKE-----TLIQQLKEELQEKNVTLDVQIQHVVEGKRA 1108
                 +   +    +Q     + E       T +Q   K   ++     T  V ++     KR
Sbjct:   531 DGSVVLDSCEVSTENDQSTRFPKPELEIQFTPLPQNKMAVKHPGCTTPVTVKIPKARKRK 590

Query:  1109 LSELTQG-VTCYKAKIKELETILETQ-KVERSHSAKLEQDILEKES 1152
             +E+ +   V C    K     T L+   +R+ S K EQ +  + S
Sbjct:   591 SNEMEEDLVKCENKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPS 636

 Score = 143 (21.5 bits), Expect = 1.3e-05, P = 1.3e-05
 Identities = 164/684 (23%), Positives = 304/684 (44%)

Query:   295 QKRKMLR-LSQDVKGYSFIKDLQWIQVSDSKEAYRLLKLGIKHQSVAFTKLNNASS---- 349
             +K  +++ L ++++ +    D+Q  V + K A   L  G+    +L
Sbjct:    49 EKETLIQQLKEELQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKV 108

Query:   350 -RSHSI-FTVKILQIEDSEMSRVIRVSELSLCDLAGSERTMKTQNEGE-RLRETGNINTS 406
              RSHS   IL+ E  +  + E  L  S +   K  N  E +L+E  T+
Sbjct:   109 ERSHSAKLEQDILEKESIILKLERNLKEFQE-HLQDSVKNTKDLNVKELKLKEEITQLTN 167

Query:   407 LLTLGKCINVLKNSEKSKFQQHVPFRESKLTHYFQSFFNGKGKICMIVNISQCYLAYDET 466
             L  K +  LK E+  +Q  + +L+  N K  + Y E
Sbjct:   168 NLQDMKHLLQLKEEEEETNRQETEKLKEELSASSARTQNLKADL---QRKEEDYADLKEK 224

Query:   467 LNVLKFSAIAQKVCVPDTLNSSQDKLFGPVKSSQDVSLDSNSNSKILNVKRATISWENSL 526
             L  K  I  Q  V  ++  +DKL  +K ++ +  N  S+ L++K+ TI
Sbjct:   225 LTDAK-KQIKQ-VQKEVSVMRDEDKLLR-IKINE-LEKKKNQCSQELDMKQRTIQQLKEQ 280

Query:   527 EDLMEDEDLVEELENAEETQNVETKLLDEDLDKTLEENKAFISHEEKRKLLDL-IEDLKK 585
             +  +  E+  +++  E  A  +  NV+ K++  ED+  TLEE  +  +  E+ ++L+  +E++++
Sbjct:   281 LNNQKVEEAIQQYERACKDLNVKEKII-EDMRMTLEEQEQ--TQVEQDQVLEAKLEEVER 337

Query:   586 KLIN-EK-KEKLT-LEFKIREEVTQEFTQYWAQREADFKETLLQEREILEE----NAERR 638
             EK KEK  LE K  +E  +  K T LQ+ E+ E  NA+R+
Sbjct:   338 LATELEKWKEKCNDLETKNNQRSNKEHEN---NTDVLGKLTNLQD-ELQESEQKYNADRK 393

Query:   639 LAIFKDLVGKCDTREEAAKDICATKVETEEATACLELKFNQIKAELAKTKGELIKTKEEL 698
             + + ++  T+ + A++I  K E ++  E  F Q + E+  ++L  +  +L
Sbjct:   394 KWLEEKMM--LITQAKEAENI-RNK-EMKKYAEDRERFFKQ-QNEMEILTAQLTEKDSDL 448

Query:   699 KKRENESDSLIQELETSNKKIITQN-QR---IKELINIIDQKEDTINEFQNLKSHMENTF 754
             +K  E D L+  LE  K +I+ N Q+  I++L  II +  +  ++K  ++
```

```
Sbjct:   449 QKWREERDQLVAALEIQLKALISSNVQKDNEIEQLKRIISETSKIETQIMDIKPKRISSA 508

Query:   755 KCNDKADTSSLIINNKLICN--ETVEVPKDSKSKICSERK---RVNENELQ-QDEP--PA 806
              DK  T L  + ++  N  E   V  DS  ++ +E     R  + EL+ Q  P  P
Sbjct:   509 D-PDKLQTEPLSTSFEISRNKIEDGSVVLDS-CEVSTENDQSTRFPKPELEIQFTPLQPN 566

Query:   807 KKGSIH--VSSAITEDQKKSEEVRPNIAEIEDIRVLQENNEGLRA---FLLTIENELKNE 861
             K    H   ++ +T   K+ + + N  E + ++   + N   R    F ++ +      +
Sbjct:   567 KMAVKHPGCTTPVTVKIPKARKRKSNEMEEDLVKCENKKNATPRTNLKFPISDDRNSSVK 626

Query:   862 KEEKAEL---NKQIVHFQQELSLSEKKNLTLSKEVQQIQSNYDIAIAELHVQKSKNQEQE 918
             KE+K  +   +K+    + + S+   NL    K+   +Q   D        + +SK ++
Sbjct:   627 KEQKVAIRPSSKKTYSLRSQASIIGV-NLATKKKEGTLQKFGDFLQHSPSILQSKAKKII 685

Query:   919 EKIM--KLSNEIETATRSITNNVSQIKLMHTKI--DELRT-LDSVSQISNID 965
             E +   KLSN +E +     NVSQ K   K+   E+ + +D  Q+  +D
Sbjct:   686 ETMSSSKLSN-VEASKE----NVSQPKRAKRKLYTSEISSPIDISGQVILMD 732

 Score = 133 (20.0 bits), Expect = 1.6e-04, P = 1.6e-04
 Identities = 94/426 (22%), Positives = 188/426 (44%)

Query:   527 EDLM-EDEDLVEELENAEETQNVETKLLDEDLDKTLEENKAFISHEEKRKLLDL-IEDLK 584
             +DL+ E E L+++L+    + +NV    LD  +    +E  +A    +        I++L+
Sbjct:    44 DDLLKEKETLIQQLKEELQEKNVT---LDVQIQHVVEGKRALSELTQGVTCYKAKIKELE 100

Query:   585 KKLINEKKEKLTLEFKIREEVTQ-EFTQYWAQREA-DFKETLLQERFILEENAERRLAIF 642
               L  +K E+ +   K+ +++ + E          +R   +F+E L    +  ++  + L +
Sbjct:   101 TILETQKVER-SHSAKLEQDILEKESIILKLERNLKEFQEHLQDSVKNTKDLNVKELKL- 158

Query:   643 KDLVGKCDTREEAAKDICATKVETEEATACLELKFNQIKAELAKTKGELIKTKEELKKRE 702
             K+ + +     +  K +    K E EE        +  ++K EL+ +      K +L++E
Sbjct:   159 KEEITQLTNNLQDMKHLLQLKEEEEETN---RQETEKLKEELSASSARTQNLKADLQRKE 215

Query:   703 NESDSLIQELETSNKKIITQNQRIKELINIIDQK-EDTINEFQNLKSHMENTFKCNDKA- 760
             +    L ++L T  KK I Q Q+    ++    D+       INE +  K+         +
Sbjct:   216 EDYADLKEKL-TDAKKQIKQVQKEVSVMRDEDKLLRIKINELEKKKNQCSQELDMKQRTI 274

Query:   761 DTSSLIINNKLICNETVE----VPKDS--KSKICSE-RKRVNENE---LQQDEPPAKKGS 810
                 +NN+ +   E ++      KD    K KI  + R  + E E  ++QD+    K
Sbjct:   275 QQLKEQLNNQKV-EEAIQQYERACKDLNVKEKIIEDMRMTLEEQEQTQVEQDQVLEAKLE 333

Query:   811 IHVSSAITEDQKKSEEVRP-NIAEIEDIRVLQENNEGLRAFLLTIENELKNEKEEKAELN 869
              V    TE +K  E+        +       ENN   +   L  +++EL+ E E+K   +
Sbjct:   334 -EVERLATELEKWKEKCNDLETKNNQRSNKEHENNTDVLGKLTNLQDELQ-ESEQKYNAD 391

Query:   870 KQIVHFQQELSLSEKKNLTLSKEVQQIQSNYDIAIAELHVQKSKNQEQEEKIMKLSNEIE 929
             ++    ++++ L    +T +KE + I++               +  K  E  E+  K  NE+E
Sbjct:   392 RK-KWLEEKMML-----ITQAKEAENIRNK----------EMKKYAEDRERFFKQQNEME 435

Query:   930 TATRSITNNVSQIKLMHTKIDEL 952
             T  +T  S ++    + D+L
Sbjct:   436 ILTAQLTEKDSDLQKWREERDQL 458
```

Pedant information for DKFZphtes3_35b4, frame 3

Report for DKFZphtes3_35b4.3

[LENGTH] 1780
[MW] 206176.77
[pI] 5.60
[HOMOL] TREMBL:U93121_1 product: "M-phase phosphoprotein-1"; Human M-phase phosphoprotein-1 mRNA, partial cds. 0.0
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YEL061c] 2e-37
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 7e-30
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 7e-30
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 3e-23
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-21

[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 6e-20
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-19
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-19
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 1e-15
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1322] 2e-14
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 2e-09
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YKL179c] 3e-09
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 2e-07
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 2e-07
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c] 2e-07
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YGL086w] 1e-06
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c] 3e-06
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YDR217c] 4e-06
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 2e-05
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YAL035w] 2e-04
[FUNCAT] r general function prediction [M. jannaschii, MJ1254] 0.001
[BLOCKS] BL00387A
[BLOCKS] BL00411H
[BLOCKS] BL00411G
[BLOCKS] BL00411F
[BLOCKS] BL00411E Kinesin motor domain proteins
[BLOCKS] BL00411D Kinesin motor domain proteins
[BLOCKS] BL00411C Kinesin motor domain proteins
[BLOCKS] BL00411B Kinesin motor domain proteins
[BLOCKS] BL00411A Kinesin motor domain proteins
[SCOP] d2kin.1 3.29.1.5.3 Kinesin [Rat (Rattus norvegicus) 2e-68
[SCOP] d2tmab_ 1.105.4.1.1 Tropomyosin [rabbit (Oryctolagus cuniculus) 4e-05
[SCOP] d3kar__ 3.29.1.5.4 Kinesin [Baker's yeast (Saccharomyce 2e-09
[EC] 3.6.1.32 Myosin ATPase 5e-25
[PIRKW] nucleus 4e-27
[PIRKW] phosphotransferase 3e-16
[PIRKW] duplication 6e-20
[PIRKW] citrulline 6e-18
[PIRKW] tandem repeat 4e-24
[PIRKW] heterodimer 3e-28
[PIRKW] endocytosis 1e-23
[PIRKW] heart 1e-17
[PIRKW] transmembrane protein 2e-28
[PIRKW] serine/threonine-specific protein kinase 3e-16
[PIRKW] zinc finger 1e-23
[PIRKW] surface antigen 2e-16
[PIRKW] DNA binding 1e-25
[PIRKW] metal binding 1e-23
[PIRKW] muscle contraction 4e-24
[PIRKW] heterotetramer 4e-24
[PIRKW] acetylated amino end 2e-19
[PIRKW] actin binding 5e-25
[PIRKW] mitosis 3e-58
[PIRKW] microtubule binding 3e-58
[PIRKW] ATP 3e-58
[PIRKW] thick filament 4e-24
[PIRKW] phosphoprotein 9e-29
[PIRKW] leucine zipper 1e-12
[PIRKW] skeletal muscle 8e-24
[PIRKW] disulfide bond 1e-12
[PIRKW] heterotrimer 1e-29
[PIRKW] calcium binding 6e-18
[PIRKW] alternative splicing 4e-21
[PIRKW] P-loop 2e-63
[PIRKW] coiled coil 3e-58
[PIRKW] heptad repeat 1e-25
[PIRKW] methylated amino acid 4e-24
[PIRKW] peripheral membrane protein 1e-23
[PIRKW] dimer 1e-12
[PIRKW] cardiac muscle 1e-17
[PIRKW] hydrolase 5e-25
[PIRKW] microtubule 6e-15
[PIRKW] muscle 7e-23
[PIRKW] membrane protein 6e-20
[PIRKW] GTP binding 8e-22
[PIRKW] EF hand 6e-18
[PIRKW] cell division 1e-25
[PIRKW] cytoskeleton 4e-24
[PIRKW] hair 6e-18
[PIRKW] Golgi apparatus 8e-24
[PIRKW] calmodulin binding 1e-23

```
[SUPFAM]        unassigned Ser/Thr or Tyr-specific protein kinases 3e-16
[SUPFAM]        myosin motor domain homology 5e-25
[SUPFAM]        alpha-actinin actin-binding domain homology 1e-13
[SUPFAM]        kinesin-related protein KIP1 9e-27
[SUPFAM]        kinesin-related protein CIN8 4e-36
[SUPFAM]        kinesin heavy chain 4e-24
[SUPFAM]        plectin 1e-13
[SUPFAM]        trichohyalin 6e-18
[SUPFAM]        kinesin-related protein KIF3 1e-29
[SUPFAM]        kinesin-related protein KIF2 3e-20
[SUPFAM]        ribosomal protein S10 homology 1e-13
[SUPFAM]        giantin 8e-24
[SUPFAM]        protein kinase homology 3e-16
[SUPFAM]        protein kinase C zinc-binding repeat homology 2e-13
[SUPFAM]        kinesin-related protein unc-104 8e-26
[SUPFAM]        human early endosome antigen 1 1e-23
[SUPFAM]        unassigned kinesin-related proteins 1e-28
[SUPFAM]        Mycoplasma genitalium hypothetical protein MG218 4e-17
[SUPFAM]        myosin heavy chain 5e-25
[SUPFAM]        conserved hypothetical P115 protein 4e-20
[SUPFAM]        centromere protein E 5e-24
[SUPFAM]        calmodulin repeat homology 6e-18
[SUPFAM]        kinesin-related protein KLP61F 1e-25
[SUPFAM]        hypothetical protein MJ0914 3e-12
[SUPFAM]        kinesin-related protein MKLP-1 2e-63
[SUPFAM]        pleckstrin repeat homology 8e-26
[SUPFAM]        hypothetical protein MJ1322 4e-13
[SUPFAM]        kinesin-related protein KIF1B 3e-28
[SUPFAM]        kinesin motor domain homology 2e-63
[SUPFAM]        kinesin-related protein KLPA 7e-25
[SUPFAM]        kinesin-related protein nodA 1e-12
[SUPFAM]        kinesin-related protein Eg5 5e-30
[PROSITE]       ATP_GTP_A       1
[PFAM]          Kinesin motor domain
[KW]            Irregular
[KW]            3D
[KW]            LOW_COMPLEXITY     7.53 %
[KW]            COILED_COIL       19.78 %

SEQ     MESNFNQEGVPRPSYVFSADPIARPSEINFDGIKLDLSHEFSLVAPNTEANSFESKDYLQ
SEG     ............................................................
COILS   ............................................................
3kar-   ............................................................

SEQ     VCLRIRPFTQSEKELESEGCVHILDSQTVVLKEPQCILGRLSEKSSGQMAQKFSFSKVFG
SEG     ............................................................
COILS   ............................................................
3kar-   ............................................................

SEQ     PATTQKEFFQGCIMQPVKDLLKGQSRLIFTYGLTNSGKTYTFQGTEENIGILPRTLNVLF
SEG     ............................................................
COILS   ............................................................
3kar-   ............................................................

SEQ     DSLQERLYTKMNLKPHRSREYLRLSSEQEKEEIASKSALLRQIKEVTVHNDSDDTLYGSL
SEG     ............................................................
COILS   ............................................................
3kar-   ............................................................

SEQ     TNSLNISEFEESIKDYEQANLNMANSIKFSVWVSFFEIYNEYIYDLFVPVSSKFQKRKML
SEG     ............................................................
COILS   ............................................................
3kar-   ............................EEEEEEEEEEETTEEEETTTCC------CCEE

SEQ     RLSQDVKGYSFIKDLQWIQVSDSKEAYRLLKLGIKHQSVAFTKLNNASSRSHSIFTVKIL
SEG     ............................................................
COILS   ............................................................
3kar-   EEETTTTE-EEEETTCCEEECCGGGHHHHHHHHHHHHCCTTTTCHHHHHHCEEEEEEEEE

SEQ     QIEDSEMSRVIRVSELSLCDLAGSERTMKTQNEGERLRETGNINTSLLTLGKCINVLKNS
SEG     ............................................................
COILS   ............................................................
3kar-   E--EETTTTCEEEEEEEEEEECCCCCCC----CCCHHHHHHHHHHHHHHHHHHHHHHHTT

SEQ     EKSKFQQHVPFRESKLTHYFQSFFNGKGKICMIVNISQCYLAYDETLNVLKFSAIAQKVC
SEG     ............................................................
COILS   ............................................................
3kar-   TTTT--TCCTTTTTHHHHHHGGGCTTTTEEEEEEEEECCCGGGHHHHHHHHHHH......

SEQ     VPDTLNSSQDKLFGPVKSSQDVSLDSNSNSKILNVKRATISWENSLEDLMEDEDLVEELE
```

```
SEG     ..........................................xxxxxxxxxxxxxxxxxx
COILS   ............................................................
3kar-   ............................................................

SEQ     NAEETQNVETKLLDEDLDKTLEENKAFISHEEKRKLLDLIEDLKKKLINEKKEKLTLEFK
SEG     xxxxxxxxxxxxxxxxxxxxxxx.......xxxxxxxxxxxxxxxxxxxxxxxxxxxx..
COILS   .............................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-   ............................................................

SEQ     IREEVTQEFTQYWAQREADFKETLLQEREILEENAERRLAIFKDLVGKCDTREEAAKDIC
SEG     ............................................................
COILS   CCCCCCC.....................................................
3kar-   ............................................................

SEQ     ATKVETEEATACLELKFNQIKAELAKTKGELIKTKEELKKRENESDSLIQELETSNKKII
SEG     ............................................................
COILS   .............................................CCCCCCCCCCCCCCC
3kar-   ............................................................

SEQ     TQNQRIKELINIIDQKEDTINEFQNLKSHMENTFKCNDKADTSSLIINNKLICNETVEVP
SEG     ............................................................
COILS   CCCCCCCCCCCCCCC.............................................
3kar-   ............................................................

SEQ     KDSKSKICSERKRVNENELQQDEPPAKKGSIHVSSAITEDQKKSEEVRPNIAEIEDIRVL
SEG     ............................................................
COILS   ........................................................CCCC
3kar-   ............................................................

SEQ     QENNEGLRAFLLTIENELKNEKEEKAELNKQIVHFQQELSLSEKKNLTLSKEVQQIQSNY
SEG     ..............xxxxxxxxxxxxxxxx..............................
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC...............
3kar-   ............................................................

SEQ     DIAIAELHVQKSKNQEQEEKIMKLSNEIETATRSITNNVSQIKLMHTKIDELRTLDSVSQ
SEG     ............................................................
COILS   .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.......................
3kar-   ............................................................

SEQ     ISNIDLLNLRDLSNGSEEDNLPNTQLDLLGNDYLVSKQVKEYRIQEPNRENSFHSSIEAI
SEG     ............................................................
COILS   ............................................................
3kar-   ............................................................

SEQ     WEECKEIVKASSKKSHQIEELEQQIEKLQAEVKGYKDENNRLKEKEHKNQDDLLKEKETL
SEG     ................xxxxxxxxxxxxxx..............................
COILS   ..........CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC............
3kar-   ............................................................

SEQ     IQQLKEELQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVERSHS
SEG     ............................................................
COILS   ......................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC....
3kar-   ............................................................

SEQ     AKLEQDILEKESIILKLERNLKEFQEHLQDSVKNTKDLNVKELKLKEEITQLTNNLQDMK
SEG     ............................................................
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....CCCCCCCCCCCCCCCCCCCCCCCCC
3kar-   ............................................................

SEQ     HLLQLKEEEEETNRQETEKLKEELSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIK
SEG     .xxxxxxxxxxxxxxxxxxx........................................
COILS   CCCCC.......................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-   ............................................................

SEQ     QVQKEVSVMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQQY
SEG     ............................................................
COILS   CCCCCCCCCCCC................................................
3kar-   ............................................................

SEQ     ERACKDLNVKEKIIEDMRMTLEEQEQTQVEQDQVLEAKLEEVERLATELEKWKEKCNDLE
SEG     ...................xxxxxxxxxxxxxxxxx........................
COILS   ..................................CCCCCCCCCCCCCCCCCCCCCCCCCC
3kar-   ............................................................

SEQ     TKNNQRSNKEHENNTDVLGKLTNLQDELQESEQKYNADRKKWLEEKMMLITQAKEAENIR
SEG     ............................................................
COILS   CC..........................................................
3kar-   ............................................................

SEQ     NKEMKKYAEDRERFFKQQNEMEILTAQLTEKDSDLQKWREERDQLVAALEIQLKALISSN
SEG     ............................................................
```

```
COILS   .......................................................
3kar-   .......................................................

SEQ     VQKDNEIEQLKRIISETSKIETQIMDIKPKRISSADPDKLQTEPLSTSFEISRNKIEDGS
SEG     ............................................................
COILS   ............................................................
3kar-   ............................................................

SEQ     VVLDSCEVSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTVEIPKARKRKSNE
SEG     ............................................................
COILS   ............................................................
3kar-   ............................................................

SEQ     MEEDLVKCENKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPSSKKTYSLRSQASIIGVN
SEG     ............................................................
COILS   ............................................................
3kar-   ............................................................

SEQ     LATKKKEGTLQKFGDFLQHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKL
SEG     ............................................................
COILS   ............................................................
3kar-   ............................................................

SEQ     YTSEISSPIDISGQVILMDQKMKESDHQIIKRRLRTKTAK
SEG     ........................................
COILS   ........................................
3kar-   ........................................
```

Prosite for DKFZphtes3_35b4.3

```
PS00017     152->160   ATP_GTP_A            PDOC00017
```

Pfam for DKFZphtes3_35b4.3

```
HMM_NAME      Kinesin motor domain

HMM               *RCRPlNeREindgcscvVQWPpWtGyktvhnghegds........phks
                   R+RP+ + E++  +  +V   + ++++   ++   +          ++
Query          64  RIRPFTQSEKELESEGCVHILDSQTVVLKEPQCILGRLSEKSSGQMAQK   112

HMM               FtFDHVFWWncTQedVYdtvAHPIVDDcFhGYNCTIFAYGQTGSGKTYTM
                  F+F +VF++++TQ++ +++ +    V+D+++G    IF+YG T SGKTYT
Query         113 FSFSKVFGPATTQKEFFQGCIMQPVKDLLKGQSRLIFTYGLTNSGKTYTF   162

HMM               MGpggehPDHmGIIPRcCHDIFdrIdkfqekDhdFW..............
                   G        +++GI+PR+++ +FD++  +   +++
Query         163 QG----TEENIGILPRTLNVLFDSLQERL-YTKMNLKPHRSREYLRLSSE   207

HMM               ..................................................

Query         208 QEKEEIASKSALLRQIKEVTVHNDSDDTLYGSLTNSLNISEFEESIKDYE   257

HMM               ............hVkCSYMEIYNEeIYDLLCPnP...qhMkpLnIHEHPN
                              +V +S++EIYNE+IYDL +P++   Q++K L++  + +
Query         258 QANLNMANSIKFSVWVSFFEIYNEYIYDLFVPVSSKFQKRKMLRLSQDVK   307

HMM               MGpYVqGCTEfHVcSYeDachWIWqGnknRHVAaTnMNdhSSRSHtIFTI
                    ++++++    V     +A +++ +G K+   VA T++N  SSRSH+IFT+
Query         308 GYSFIKDLQWIQVSDSKEAYRLLKLGIKHQSVAFTKLNNASSRSHSIFTV   357

HMM               HVeQrHk.qcdehvcHSKMNLVDLAGSERvnrTGAEGQRlKEGcNINqSL
                  ++ Q +     +  +++S ++L DLAGSER+ +T+ EG RL+E +NIN SL
Query         358 KILQIEDSEMSRVIRVSELSLCDLAGSERTMKTQNEGERLRETGNINTSL   407

HMM               ttLGnVInaLaDgqTKYmYgghgHIPYRDSKLTWlLQDSLGGNcKTcMIA
                  +TLG++IN+L    +    + +  +H+P+R+SKLT+ +Q  + G +K CMI+
Query         408 LTLGKCINVLKNSE---KSKFQQHVPFRESKLTHYFQSFFNGKGKICMIV   454

HMM               CIWPadWNYEETLSTLRYAdRAKnIkNkPQINEDPca*
                  +I+    + Y+ETL++L++ + A+++ +   ++N+++++
Query         455 NISQCYLAYDETLNVLKFSAIAQKVCVPDTLNSSQDK   491
```

DKFZphtes3_35b5

----

group: metabolism

DKFZphtes3_35b5 encodes a novel 466 amino acid protein, with similarity to bovine accessory subunit for vacuolar ATPase and rat C7-1 protein.

The vacuolar proton-ATPase (V-ATPase) translocates protons into intracellular organelles or across the plasma membrane of specialized cells. The catalytic domain consists of a hexamer of 3 A subunits and 3 B subunits, plus accessory subunits C, D, and E. The rat homolog C7-1 seems to be enriched in aged adult rats in the frontal cortex.

The novel protein can find application in modulating the v-ATPase activity in endocytic and secretory organelles.

strong similarity to bovine vacuolar ATPase (EC 3.6.1.-) chain A

complete cDNA, complete cds potential start at Bp 8 , EST hits matches perfect to I54197 hypothetical protein, but posess 186 aa additional at N-terminus

Sequenced by DKFZ

Locus: unknown

Insert length: 2043 bp
Poly A stretch at pos. 2033, polyadenylation signal at pos. 2012

```
   1 GGCGGCCATG GCGACGGCTC GAGTGCGGAT GGGGCCGCGG TGCGCCCAGG
  51 CGCTCTGGCG CATGCCGTGG CTGCCGGTGT TTTTGTCGTT GGCGGCGGCG
 101 GCGGCGGCGG CAGCGGCGGA GCAGCAGGTC CCGCTGGTGC TGTGGTCGAG
 151 TGACCGGGAC TTGTGGGCTC CTGCGGCCGA CACTCATGAA GGCCACATCA
 201 CCAGCGACTT GCAGCTCTCT ACCTACTTAG ATCCCGCCCT GGAGCTGGGT
 251 CCCAGGAATG TGCTGCTGTT CCTGCAGGAC AAGCTGAGCA TTGAGGATTT
 301 CACAGCATAT GGCGGTGTGT TTGGAAACAA GCAGGACAGC GCCTTTTCTA
 351 ACCTAGAGAA TGCCCTGGAC CTGGCCCCCT CCTCACTGGT GCTTCCTGCC
 401 GTCGACTGGT ATGCAGTCAG CACTCTGACC ACTTACCTGC AGGAGAAGCT
 451 CGGGGCCAGC CCCTTGCATG TGGACCTGGC CACCCTGCGG GAGCTGAAGC
 501 TCAATGCCAG CCTCCCTGCT CTGCTGCTCA TTCGCCTGCC CTACACAGCC
 551 AGCTCTGGTC TGATGGCACC CAGGGAAGTC CTCACAGGCA ACGATGAGGT
 601 CATCGGGCAG GTCCTGAGCA CACTCAAGTC CGAAGATGTC CCATACACAG
 651 CGGCCCTCAC AGCGGTCCGC CCTTCCAGGG TGGCCCGTGA TGTAGCCGTG
 701 GTGGCCGGAG GGCTAGGTCG CCAGCTGCTA CAAAAACAGC CAGTATCACC
 751 TGTGATCCAT CCTCCTGTGA GTTACAATGA CACCGCTCCC CGGATCCTGT
 801 TCTGGGCCCA AAACTTCTCT GTGGCGTACA AGGACCAGTG GGAGGACCTG
 851 ACTCCCCTCA CCTTTGGGGT GCAGGAACTC AACCTGACTG GCTCCTTCTG
 901 GAATGACTCC TTTGCCAGGC TCTCACTGAC CTATGAACGA CTCTTTGGTA
 951 CCACAGTGAC ATTCAAGTTC ATTCTGGCCA ACCGCCTCTA CCCAGTGTCT
1001 GCCCGGCACT GGTTTACCAT GGAGCGCCTC GAAGTCCACA GCAATGGCTC
1051 CGTCGCCTAC TTCAATGCTT CCCAGGTCAC AGGGCCCAGC ATCTACTCCT
1101 TCCACTGCGA GTATGTCAGC AGCCTGAGCA AGAAGGGTAG TCTCCTCGTG
1151 GCCCGCACGC AGCCCTCTCC CTGGCAGATG ATGCTTCAGG ACTTCCAGAT
1201 CCAGGCTTTC AACGTAATGG GGGAGCAGTT CTCCTACGCC AGCGACTGTG
1251 CCAGCTTCTT CTCCCCCGGC ATCTGGATGG GGCTGCTCAC CTCCCTGTTC
1301 ATGCTCTTCA TCTTCACCTA TGGCCTGCAC ATGATCCTCA GCCTCAAGAC
1351 CATGGATCGC TTTGATGACC ACAAGGGCCC CACTATTTCT TTGACCCAGA
1401 TTGTGTGACC CTGTGCCAGT GGGGGGGTTG AGGGTGGGAC GGTGTCCGTG
1451 TTGTTGCTTT CCCACCCTGC AGCGCACTGG ACTGAAGAGC TTCCCTCTTC
1501 CTACTGCAGC ATGAACTGCA AGCTCCCCTC AGCCCATCTT GCTCCCTCTT
1551 CAGCCCGCTG AGGAGCTTTC TTGGGCTGCC CCCATCTCTC CCAACAAGGT
1601 GTACATATTC TGCGTAGATG CTAGACCAAC CAGCTTCCCA GGGTTCGTCG
1651 CTGTGAGGCG TAAGGGACAT GAATTCTAGG GTCTCCTTTC TCCTTATTTA
1701 TTCTTGTGGC TACATCATCC CTGGCTGTGG ATAGTGCTTT TGTGTAGCAA
1751 ATGCTCCCTC CTTAAGGTTA TAGGGCTCCC TGAGTTTGGG AGTGTGGAAG
1801 TACTACTTAA CTGTCTGTCC TGCTTGGCTG CCGTTATCGT TTTCTGGTGA
1851 TGTTGTGCTA ACAATAAGAA GTACACGGGT TTATTTCTGT GGCCTGAGAA
1901 GGAAGGGACC TCCACGACAG GTGGGCTGGG TGCGATCGCC GGCTGTTTGG
1951 CATGTTCCCA CCGGGAGTGC CGGGCAGGAG CATGGGGTGC TTGGTTGTTT
2001 CCTTCCTAAT AAAATAAACG CGGGTCGCCA TGCAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

95014142:
A novel accessory subunit for vacuolar H(+)-ATPase from chromaffin granules.

97215246:
Identification of a rat brain gene associated with aging by PCR differential display method.

Peptide information for frame 2

ORF from 8 bp to 1405 bp; peptide length: 466
Category: strong similarity to known protein

```
  1 MATARVRMGP RCAQALWRMP WLPVFLSLAA AAAAAAAEQQ VPLVLWSSDR
 51 DLWAPAADTH EGHITSDLQL STYLDPALEL GPRNVLLFLQ DKLSIEDFTA
101 YGGVFGNKQD SAFSNLENAL DLAPSSLVLP AVDWYAVSTL TTYLQEKLGA
151 SPLHVDLATL RELKLNASLP ALLLIRLPYT ASSGLMAPRE VLTGNDEVIG
201 QVLSTLKSED VPYTAALTAV RPSRVARDVA VVAGGLGRQL LQKQPVSPVI
251 HPPVSYNDTA PRILFWAQNF SVAYKDQWED LTPLTFGVQE LNLTGSFWND
301 SFARLSLTYE RLFGTTVTFK FILANRLYPV SARHWFTMER LEVHSNGSVA
351 YFNASQVTGP SIYSFHCEYV SSLSKKGSLL VARTQPSPWQ MMLQDFQIQA
401 FNVMGEQFSY ASDCASFFSP GIWMGLLTSL FMLFIFTYGL HMILSLKTMD
451 RFDDHKGPTI SLTQIV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35b5, frame 2

TREMBL:AF035387_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1 protein (C7-1) mRNA, complete cds., N = 1, Score = 2088, P = 3.8e-216

PIR:A55116 vacuolar ATPase (EC 3.6.1.-) chain Ac45 - bovine, N = 1, Score = 2011, P = 5.5e-208

PIR:I54197 hypothetical protein - human, N = 1, Score = 1464, P = 5.1e-150

```
>TREMBL:AF035387_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus
     C7-1 protein (C7-1) mRNA, complete cds.
          Length = 463

 HSPs:

 Score = 2088 (313.3 bits), Expect = 3.8e-216, P = 3.8e-216
 Identities = 408/463 (88%), Positives = 426/463 (92%)

Query:     4 ARVRMGPRCAQALWRMPWLPVFLSLAAAAAAAAAEQQVPLVLWSSDRDLWAPAADTHEGH 63
             +R+R G R A  LW      + LSL A AAA AAEQQVPLVLWSSDRDLWAP ADTHEGH
Sbjct:     8 SRIRTGTRWAPVLW------LLLSLVAVAAAVAAEQQVPLVLWSSDRDLWAPVADTHEGH 61

Query:    64 ITSDLQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA 123
             ITSD+QLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA
Sbjct:    62 ITSDMQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA 121

Query:   124 PSSLVLPAVDWYAVSTLTTYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYTASS 183
             PSSLVLPAVDWYA+STLTTYLQEKLGASPLHVDLATL+ELKLNASLPALLLIRLPYTASS
Sbjct:   122 PSSLVLPAVDWYAISTLTTYLQEKLGASPLHVDLATLKELKLNASLPALLLIRLPYTASS 181

Query:   184 GLMAPREVLTGNDEVIGQVLSTLKSEDVPYTAALTAVRPSRVARDVAVVAGGLGRQLLQK 243
             GLMAPREVLTGNDEVIGQVLSTL+SEDVPYTAALTAVRPSRVARDVA+VAGGLGRQLLQ
Sbjct:   182 GLMAPREVLTGNDEVIGQVLSTLESEDVPYTAALTAVRPSRVARDVAMVAGGLGRQLLQT 241

Query:   244 QPVSPVIHPPVSYNDTAPRILFWAQNFSVAYKDQWEDLTPLTFGVQELNLTGSFWNDSFA 303
             Q  SP IHPPVSYNDTAPRILFWAQNFSVAYKD+W+DLT LTFGV+ LNLTGSFWNDSFA
Sbjct:   242 QVASPAIHPPVSYNDTAPRILFWAQNFSVAYKDEWKDLTSLTFGVENLNLTGSFWNDSFA 301

Query:   304 RLSLTYERLFGTTVTFKFILANRLYPVSARHWFTMERLEVHSNGSVAYFNASQVTGPSIY 363
              LSLTYE LFG TVTFKFILA+R YPVSAR+WFTMERLE+HSNGSVA+FN SQVTGPSIY
```

```
Sbjct:  302 MLSLTYEPLFGATVTFKFILASRFYPVSARYWFTMERLEIHSNGSVAHFNVSQVTGPSIY 361

Query:  364 SFHCEYVSSLSKKGSLLVARTQPSPWQMMLQDFQIQAFNVMGEQFSYASDCASFFSPGIW 423
            SFHCEYVSSLSKKGSLLV    PS WQM L +FQIQAFNV GEQFSYASDCA FFSPGIW
Sbjct:  362 SFHCEYVSSLSKKGSLLVTNV-PSLWQMTLHNFQIQAFNVTGEQFSYASDCAGFFSPGIW 420

Query:  424 MGLLTSLFMLFIFTYGLHMILSLKTMDRFDDHKGPTISLTQIV 466
            MGLLT+LFMLFIFTYGLHMILSLKTMDRFDD KGPTI+LTQIV
Sbjct:  421 MGLLTTLFMLFIFTYGLHMILSLKTMDRFDDRKGPTITLTQIV 463
```

Pedant information for DKFZphtes3_35b5, frame 2

Report for DKFZphtes3_35b5.2

```
[LENGTH]        466
[MW]            51621.44
[pI]            5.73
[HOMOL]         TREMBL:AF035387_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1
protein (C7-1) mRNA, complete cds. 0.0
[PIRKW]         hydrolase 0.0
[PROSITE]       MYRISTYL        7
[PROSITE]       CAMP_PHOSPHO_SITE      1
[PROSITE]       CK2_PHOSPHO_SITE       7
[PROSITE]       TYR_PHOSPHO_SITE       1
[PROSITE]       PKC_PHOSPHO_SITE       8
[PROSITE]       ASN_GLYCOSYLATION      7
[KW]            SIGNAL_PEPTIDE 38
[KW]            TRANSMEMBRANE  1
[KW]            LOW_COMPLEXITY    11.59 %

SEQ   MATARVRMGPRCAQALWRMPWLPVFLSLAAAAAAAAAEQQVPLVLWSSDRDLWAPAADTH
SEG   ............................xxxxxxxxx.......................
PRD   ccceeeeccccchhhhhhhccccchhhhhhhhhhhhhhhhhcccceeeeccccccccccccc
MEM   ............................................................

SEQ   EGHITSDLQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENAL
SEG   ............................................................
PRD   cccccchhhhhccccccccccccceeeccccccccccccccccccccchhhhhhhhcc
MEM   ............................................................

SEQ   DLAPSSLVLPAVDWYAVSTLTTYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYT
SEG   .........................................xxxxxxxxxxxxxxx...
PRD   ccccccccccccceeeeehhhhhhhhhhccccchhhhhhhhhhhhhcchhhhhhhcccc
MEM   ............................................................

SEQ   ASSGLMAPREVLTGNDEVIGQVLSTLKSEDVPYTAALTAVRPSRVARDVAVVAGGLGRQL
SEG   .....................................xxxxxxxxxxxxxxxxxxxx..
PRD   ccccccceeeeeeccccchhhhhhhcccccchhhhhhhccccceeehhhhhcccchhh
MEM   ............................................................

SEQ   LQKQPVSPVIHPPVSYNDTAPRILFWAQNFSVAYKDQWEDLTPLTFGVQELNLTGSFWND
SEG   ............................................................
PRD   hhhhcccccccccccccccccceeeeecccccceeecccccccccceeeeeecccccccccc
MEM   ............................................................

SEQ   SFARLSLTYERLFGTTVTFKFILANRLYPVSARHWFTMERLEVHSNGSVAYFNASQVTGP
SEG   ............................................................
PRD   hhhhhhhhhhhhccceeeeeeeecccccccccchhhhhhhhhccccceeeeeeeccccc
MEM   ............................................................

SEQ   SIYSFHCEYVSSLSKKGSLLVARTQPSPWQMMLQDFQIQAFNVMGEQFSYASDCASFFSP
SEG   ..........xxxxxxxxxx........................................
PRD   ceeeeeeeeeeccccceeeeeccccchhhhhhhheeeeccccccccccccccccccc
MEM   ......................................................MMMMMM

SEQ   GIWMGLLTSLFMLFIFTYGLHMILSLKTMDRFDDHKGPTISLTQIV
SEG   ..............................................
PRD   ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccceeeeccc
MEM   MMMMMMMMMMMMMMMMMMMMMMM.......................
```

Prosite for DKFZphtes3_35b5.2

PS00001	166->170	ASN_GLYCOSYLATION	PDOC00001
PS00001	257->261	ASN_GLYCOSYLATION	PDOC00001
PS00001	269->273	ASN_GLYCOSYLATION	PDOC00001

PS00001	292->296	ASN_GLYCOSYLATION	PDOC00001
PS00001	299->303	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00001	353->357	ASN_GLYCOSYLATION	PDOC00001
PS00004	375->379	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	159->162	PKC_PHOSPHO_SITE	PDOC00005
PS00005	205->208	PKC_PHOSPHO_SITE	PDOC00005
PS00005	318->321	PKC_PHOSPHO_SITE	PDOC00005
PS00005	331->334	PKC_PHOSPHO_SITE	PDOC00005
PS00005	374->377	PKC_PHOSPHO_SITE	PDOC00005
PS00005	445->448	PKC_PHOSPHO_SITE	PDOC00005
PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	114->118	CK2_PHOSPHO_SITE	PDOC00006
PS00006	159->163	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00007	207->214	TYR_PHOSPHO_SITE	PDOC00007
PS00008	102->108	MYRISTYL	PDOC00008
PS00008	103->109	MYRISTYL	PDOC00008
PS00008	200->206	MYRISTYL	PDOC00008
PS00008	295->301	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	421->427	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35b5.2)

DKFZphtes3_35e21
----

group: differentiation/development

DKFZphtes3_35e21.2 encodes a novel 104 amino acid putative interleukin precursor, related to interleukin-7.

Due to the close relationship to human interleukin-7, the novel interleukin is expected to act as a new growth factor for human B lineage cells. Additionally, the protein should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells.

This new interleukin could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells.

similarity to interleukin-7 precursor

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 2095 bp
Poly A stretch at pos. 2085, polyadenylation signal at pos. 2067

```
   1 GGATGAAAGT GATTTAATTC ATTTTTAGAA TTTTTTTTTT GTTTTGTTTT
  51 AGCAACATGC TGAACAACTA ATTTACTTTA AAAATAAGCC AGTTAAAACA
 101 AAGGACGCTA AGCCCAAGTG GGGGGCAATA TTAGTCAGGA TCTTTGGGGT
 151 CTAATTCCAG ACCAACTTTC AGAAGCACTT CTTTGTCTCT GTTCTCACCT
 201 CTGCTGTCCC TCTCTTCCCT CATCCCCTAA GAGAGACAAA GATAAAAGCC
 251 CACCTGCATC CCTAAGTCTT ACTGAGATCA GCCACCCCAG GGGAGAGAAA
 301 CTGGATCTAC TTACAGCCAC CCCCTGTTTC CATCCATATA CTTACTTCCC
 351 CCAATTTGCA TGTGATTATG GAAACAAGTC ATGCTCATGA AAGCAACTGT
 401 AAAATAAAAG GTTATGGAGT AGTTCAGCAA CTTCTTCACA GCCAGCTTTG
 451 TGGAGCTGGG GAGGACTTAG GGCCCATTGG AGTCTCTTAT GTGTACAGCT
 501 TCAGGGCTGT CCCTTTCAGT TTGATTTTAA GCAATGCCTC ACTTCATAGC
 551 TTAGGGGGTA AGGATTCCAT TCAGGTAGGT TGTCTAAAGG AACTAATGGG
 601 ACCTCTCAGT GAATTAGCTG ACCAGATTTT AGGAAATCTT TTTAATTTCT
 651 ATGATTTTCC TTCTCACATT TTGAAATGGT AAAATTGACT GGAAATAATT
 701 TTTCTTGGTG CCTTATTGGT TTTCCTTGCA AACCTTTCTC ATATTTTCTC
 751 ATGACCATTG CCAGTGACCA AGGCCCATGT GTGTGTTGTG TGTAATTGTG
 801 GGCATGTACA AGCTTAAATA ACGTGCCGAC AGCACTGTTT CAAAGTTGGT
 851 ATTCATTAGG CTGTTGCCTC CTGGGCTGGA GCTGCGCTAA TCCTGACACC
 901 GGCTGCCAGG AGAAAACCTC ATGGATCACA CACCAAACCT TAATAACAGC
 951 ATCCGTGACC TGCACTCTCC AGTACAGAAT GGGAACCCCA GAGCTAGGAA
1001 ATGTAGTTGT ATATTTTAAT GAACTGCTAC CCCAGCCAAA GAAGCTTCTT
1051 TCACTTTTGT GCTCTACAGA AAGCCCAAGG GGGGTAGGAG GGACAGAGCT
1101 TTGAATAACT GCTTTCTAAC ACTAAATGTG GCCAACAGGA CAGAGCACAT
1151 CACACGTATA GGCAGGTGTG AGGGACAGTG GCTAAGAATT GCCTGCTCCC
1201 TCTGCATGCT CTTTCTTGTT TCCAAAGTCC AATCAAGTGA TCCTGGGAAA
1251 CAAATCTGTC TGGATTGCGG AGGGTGGTTC TGAAAGAACT GCCAAGACGT
1301 TAAAGAAGGG TGAAGAGTAG GCAGAATATA AGTAGCTAAC CTGAGTCAAG
1351 ACTCTCAAAA GCTAGCAGCC TGATGACAAT AGGATTTATT TCAGCCAGGA
1401 TAGTGTCTGT CTGTGAGTGC ATCATTTTAA GACAGTATGA CTTCATGTTG
1451 TTACAAACTA TGTATAGTAT GTATGTTTTG TGGGTTGTAT ATATACATAA
1501 TATATATTAT ATATATATAT GAGAGATTTG GTGACTTTTG ATACGGGTTT
1551 GGTGCAGGTG AATTTATTAC TGAGCCAAAT GAGGCACATA CCGAGTCAGT
1601 AGTTGAAGTC CAGGGCATTC GATACTGTTT ATGATTTCCA TATATGTATA
1651 GTGCCTATCC CATGCTGTAG TCACTGTTAT GTTAAATCCA GAAGTTACAC
1701 TAGAGCCAGC GATACTTTAT TTGTAGACAA TCAATTTGAA TCCATATGTT
1751 ATTACTGGCA GATGATACAT GATTACAGTT CTGAATCTGT AACACTTACA
1801 AAAGGAAACC CAGAGCAGCT TGATGAGTTT TTGTTTCTGC TTCGTTCCTG
1851 GGAGTCAGTA GAAACAGCAG TTGTATGTGG TTATGTTAGT CTCAAGATAC
1901 TTAATTTGTT GACCTTACTT CAGAAAAATT TTGTATGTAT TATATTTGTG
1951 GGAAGGTAAA ATAATCATTT GAGATTTTTA TCAAATATGA AGATTAGTTA
2001 TTTATGAAAA ACAAAGAAAT GTCTATTTTT CTTTGTTCCC AATTAATGTA
2051 GATAAATTTT AAAATGCATT AAAGTAATGG TCCGGAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

89098903:
Human interleukin 7: molecular cloning and growth factor activity on human and murine B-lineage cells.

Peptide information for frame 2

ORF from 368 bp to 679 bp; peptide length: 104
Category: similarity to known protein

```
  1 METSHAHESN CKIKGYGVVQ QLLHSQLCGA GEDLGPIGVS YVYSFRAVPF
 51 SLILSNASLH SLGGKDSIQV GCLKELMGPL SELADQILGN LFNFYDFPSH
101 ILKW
```

BLASTP hits

Entry B32223 from database PIR:
interleukin-7 precursor (clone 1) - human
Score = 66, P = 7.0e-01, identities = 21/70, positives = 33/70

Alert BLASTP hits for DKFZphtes3_35e21, frame 2

PIR:B32223 interleukin-7 precursor (clone 1) - human, N = 1, Score = 66, P = 0.72

TREMBL:PADAL1_1 gene: "dal1"; P.abies dal1 mRNA, N = 2, Score = 59, P = 0.77

PIR:C32223 interleukin-7 precursor (clone 4) - human, N = 1, Score = 66, P = 0.79

TREMBL:PRU76726_1 gene: "PrMADS3"; product: "MADS-box protein"; Pinus radiata MADS-box protein (PrMADS3) mRNA, complete cds., N = 2, Score = 59, P = 0.94

>PIR:B32223 interleukin-7 precursor (clone 1) - human
Length = 133

HSPs:

Score = 66 (9.9 bits), Expect = 1.3e+00, P = 7.2e-01
Identities = 21/68 (30%), Positives = 33/68 (48%)

```
Query:  39 VSYVYSFRAVPFSLIL-----SNASLHSLGGK--DSIQVGCLKELMGPLSELADQILGNL 91
           VS+ Y F   P  L+L      S+  +    GK  +S+ +  + +L+  + E+     L N
Sbjct:   4 VSFRYIFGLPPLILVLLPVASSDCDIEGKDGKQYESVLMVSIDQLLDSMKEIGSNCLNNE 63

Query:  92 FNFYDFPSHI 101
           FNF  F  HI
Sbjct:  64 FNF--FKRHI 71
```

Pedant information for DKFZphtes3_35e21, frame 2

Report for DKFZphtes3_35e21.2

[LENGTH]	104	
[MW]	11339.12	
[pI]	5.87	
[PROSITE]	MYRISTYL	2
[PROSITE]	PKC_PHOSPHO_SITE	1
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	Alpha_Beta	

```
SEQ  METSHAHESNCKIKGYGVVQQLLHSQLCGAGEDLGPIGVSYVYSFRAVPFSLILSNASLH
PRD  ccchhhhhcccccccchhhhhhhhhhcccccccccceeeeeeecccccceeeeecccccc
```

```
SEQ     SLGGKDSIQVGCLKELMGPLSELADQILGNLFNFYDFPSHILKW
PRD     cccccceeeccccccccccchhhhhhhhcccccccccccccccc
```

Prosite for DKFZphtes3_35e21.2

PS00001	56->60	ASN_GLYCOSYLATION	PDOC00001
PS00005	44->47	PKC_PHOSPHO_SITE	PDOC00005
PS00008	63->69	MYRISTYL	PDOC00008
PS00008	89->95	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35e21.2)

DKFZphtes3_35g6
----

group: testes derived

DKFZphtes3_35g6 encodes a novel 482 amino acid protein with high partial similarity to H. sapiens chromosome 19, cosmid R27216.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to R27216_1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="15"

Insert length: 3177 bp
Poly A stretch at pos. 3167, polyadenylation signal at pos. 3148

```
   1 GGAGGCAGCG CCGGCCTCCG GAGGCGGCCT GGGCGATGGC GGCGGAGTTT
  51 TGTCCATAAC CTGGGCAACC GCGCAGCTGG AGGATGGCCT CACTCGGGCC
 101 TGCCGCAGCT GGGGAGCAGG CGTCGGGGGC TGAGGCGGAG CCGGGCCCCG
 151 CGGGGCCGCC GCCGCCGCCC TCACCGTCCT CTCTGGGGCC CCTGCTCCCC
 201 CTGCAGCGGG AACCTCTCTA CAACTGGCAG GCGACCAAGG CGTCGCTGAA
 251 GGAGCGCTTC GCCTTCCTCT TCAACTCGGA GCTGCTGAGC GATGTGCGCT
 301 TCGTACTGGG CAAGGGTCGC GGCGCCGCCG CCGCTGGGGG CCCGCAGCGC
 351 ATCCCCGCCC ACCGCTTCGT GCTGGCGGCC GGCAGCGCCG TCTTTGACGC
 401 CATGTTCAAC GGCGGCATGG CCACCACGTC GGCCGAGATC GAGCTGCCGG
 451 ACGTGGAGCC CGCAGCCTTC CTGGCGCTGC TGAGATTTCT ATATTCAGAT
 501 GAAGTTCAAA TTGGTCCAGA AACAGTTATG ACCACTCTTT ATACTGCCAA
 551 GAAATACGCA GTCCCAGCCT TGGAAGCACA CTGTGTAGAA TTTCTCACCA
 601 AACATCTTAG GGCAGATAAT GCCTTTATGT TACTTACTCA GGCTCGATTA
 651 TTTGATGAAC CTCAGCTTGC TAGTCTTTGT CTAGATACAA TAGACAAAAG
 701 CACAATGGAT GCAATAAGTG CAGAAGGGTT TACTGATATT GATATAGATA
 751 CACTCTGTGC AGTTTTAGAG AGAGACACAC TCAGTATTCG AGAAAGTCGA
 801 CTTTTTGGAG CTGTTGTACG CTGGGCAGAA GCAGAATGTC AGAGACAACA
 851 ATTACCTGTG ACTTTTGGGA ATAAACAAAA AGTTCTAGGA AAAGCACTTT
 901 CCTTAATCCG GTTCCCACTG ATGACAATTG AGGAATTTGC AGCAGGTCCT
 951 GCTCAATCTG GAATTTTGTC AGATCGTGAA GTGGTAAACC TCTTTCTTCA
1001 TTTTACTGTC AACCCTAAAC CCCGAGTTGA ATACATTGAC CGACCAAGAT
1051 GCTGTCTCAG GGGAAAGGAA TGCTGCATCA ATAGATTCCA GCAAGTAGAA
1101 AGCCGCTGGG GTTACAGTGG GACGAGTGAT CGAATCAGAT TCACAGTTAA
1151 TAGAAGGATC TCTATAGTTG GATTTGGCTT GTATGGATCT ATTCATGGCC
1201 CTACAGATTA TCAAGTGAAT ATACAGATCA TTGAATATGA GAAAAAGCAA
1251 ACCCTGGGAC AGAATGATAC CGGCTTTAGT TGTGATGGGA CAGCTAACAC
1301 ATTCAGGGTC ATGTTCAAGG AACCCATAGA GATCCTGCCC AATGTGTGCT
1351 ACACAGCATG TGCAACACTC AAAGGTCCAG ATTCCCACTA TGGCACAAAA
1401 GGATTGAAGA AAGTAGTGCA TGAGACACCT GCTGCAAGCA AGACTGTTTT
1451 TTTCTTTTTT AGTTCCCCTG GCAATAATAA TGGCACTTCA ATAGAAGATG
1501 GACAAATTCC AGAAATCATA TTTTATACAT AATTTAGCAT TATAATACAT
1551 CTTGGCTAAA TAATACCATA CAATCTAGTG TCAAAAACAT AAATGGCCAC
1601 AAAAAAGTAG TTTGAGTGTT ATGAATATTT AAAATTGTAA GATAAGAAAC
1651 AGTTTCTTAG AGCAGATAGA AAAATGCTTA TTTAAATCTT TGCATGATTT
1701 AAAAACAGAT TTTCCATTTT CTTACAACTT TAAGAGAAAA GAACTGGGTT
1751 TAATGGTTTA AAAAAAAGCA CAGCTTTTTC ACCTTCATCT TGTATAATTT
1801 CATAGATTGG CTGACTTAGG GTCTTTCAAT AGTTTGGGAA TTGAAAGATT
1851 CTTGTTATAT ATAGCTAGTT TGGGTTTGTT TTTGTTTTAA CTATTTTGAA
1901 GGTTAGGTGA GATGGGCAAA TAGGCTTAAC TATTTTGAAG GTTGGATGAA
1951 AAGAGATGGG TCAGTATTCC TACAGAATTC TTATTAACTC AAATAACTAA
2001 ATTTCAGAAA ATTAAGAAGC TGACTTTATA TTTGGTGGTT TGAAGTATCT
2051 TGTTGTTAGC ATTTGTAATA ATGCTAAAAA AGGCCTAATA AAATGCCCAA
2101 GAAAATATTC AGTGCATTTA TAGAGAAGGA TATTTTGTAG TAGTATAGTA
2151 ATGTGTTATG TAGTACAGTT TTAAAGCTAT AAATGGAATT TTGTGTAAAT
2201 TCACAAAAAT GTGATATAAA CAGGATCTAA GACTGGATTC CCTGTCACTA
2251 AACTGCACCA CTATACCTGT CTCTCTGTGT GGGGGACACT GCTGATGATT
2301 CCCAAGATTG AGATGATGAC GGTGATGACG ACTGGGTGAA CAGCCATCAC
2351 TTCAACATTG TGATAATCCT TCACAGCAAG AAACCGAATA AAATACTAAC
2401 ATTTCTAACA ACTGCTCTGA CATTGTAAAG AGATCCAACA GAATCACTCC
2451 TGCTGAAAAA TACGCTTTCT GCCACCTACA CATTTCTATT TAGGAAGTAA
2501 AATTTGCTTC ATGGTCATGA CCCCATTAGT CAGTGTTACA GCTGTGTTGG
2551 GGATAGGAAG TATATCTGGC AGATTGACAT TTATACACTT TTTTATAAAG
2601 CAGATTTTAA AATATAGTAA CATCCATTTT TTTCCCTTGA AAGTGATTCT
2651 CTTATAAAAA ATGAAAGTGG AGTTTAAGGT ATATCAAATC GTTGTGGAAG
2701 GTGATTAAAA ATCAAAATTC TTTTAAATAT CAACTTAATT TTTTCTAAGT
```

```
2751 AAGATACAAA AAATTTTCAT CTAAAGTAAT ATTTCACTTT ATATTGTAAA
2801 GAAGGTAGGT ATATTGGTGG CTGAGGTCTC TTGAAATTGC TAAAGGGAAA
2851 TTTTTCTATG GTAATGCTCT TACGGATATA AGCCTCAGTT AAATGGAATT
2901 ATCTATGGGA TGTGTGGTTC TGGTTAACTA AAAATTAACC AGTAAACACT
2951 CTGTAGTAAC CATTACAGAA AATACTTCTG CCTTAAAAAA TATGATATGC
3001 CAGAGATGAG TTAGTGTTTC TTGACGTTGG AGACCTATAA ATGCCTCATC
3051 TGTTGTACTG AACAATTGAA ACTGCATGCA GCCATAAAAG GGACAAGAAA
3101 CAGAACTGTT TACTAACTTT GGGACATCCC CTGGAGTTTT TAAAAATAAA
3151 TAAATATATA TATATATAAA AAAAAAA
```

BLAST Results

Entry G37753 from database EMBL:
SHGC-63477 Human Homo sapiens STS genomic.
Score = 1627, P = 3.0e-66, identities = 327/329

Entry G37752 from database EMBL:
SHGC-63476 Human Homo sapiens STS genomic.
Score = 1578, P = 6.2e-64, identities = 320/324

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 84 bp to 1529 bp; peptide length: 482
Category: similarity to unknown protein

```
  1 MASLGPAAAG EQASGAEAEP GPAGPPPPPS PSSLGPLLPL QREPLYNWQA
 51 TKASLKERFA FLFNSELLSD VRFVLGKGRG AAAAGGPQRI PAHRFVLAAG
101 SAVFDAMFNG GMATTSAEIE LPDVEPAAFL ALLRFLYSDE VQIGPETVMT
151 TLYTAKKYAV PALEAHCVEF LTKHLRADNA FMLLTQARLF DEPQLASLCL
201 DTIDKSTMDA ISAEGFTDID IDTLCAVLER DTLSIRESRL FGAVVRWAEA
251 ECQRQQLPVT FGNKQKVLGK ALSLIRFPLM TIEEFAAGPA QSGILSDREV
301 VNLFLHFTVN PKPRVEYIDR PRCCLRGKEC CINRFQQVES RWGYSGTSDR
351 IRFTVNRRIS IVGFGLYGSI HGPTDYQVNI QIIEYEKKQT LGQNDTGFSC
401 DGTANTFRVM FKEPIEILPN VCYTACATLK GPDSHYGTKG LKKVVHETPA
451 ASKTVFFFFS SPGNNNGTSI EDGQIPEIIF YT
```

BLASTP hits

Entry AC005306_2 from database TREMBL:
product: "R27216_1"; Homo sapiens chromosome 19, cosmid R27216, complete sequence.
Score = 1298, P = 1.9e-132, identities = 245/297, positives = 268/297

Entry CEF38H4_9 from database TREMBLNEW:
gene: "F38H4.7"; Caenorhabditis elegans cosmid F38H4
Score = 1237, P = 5.6e-126, identities = 248/446, positives = 322/446

Entry AC004678_1 from database TREMBL:
product: "R34094_1"; Homo sapiens chromosome 19, cosmid R34094, complete sequence.
Score = 555, P = 1.0e-53, identities = 112/137, positives = 123/137

Alert BLASTP hits for DKFZphtes3_35g6, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_35g6, frame 3

Report for DKFZphtes3_35g6.3

[LENGTH]	482
[MW]	52771.47
[pI]	5.79

```
[HOMOL]         TREMBL:AC005306_2 product: "R27216_1"; Homo sapiens chromosome 19, cosmid
R27216, complete sequence. 1e-142
[BLOCKS]        BL01075D Acetate and butyrate kinases family proteins
[SUPFAM]        POZ domain homology 3e-08
[SUPFAM]        A55R protein middle region homology 5e-06
[SUPFAM]        A55R protein 5e-06
[SUPFAM]        A55R protein carboxyl-terminal homology 5e-06
[PROSITE]       MYRISTYL        6
[PROSITE]       CAMP_PHOSPHO_SITE       2
[PROSITE]       CK2_PHOSPHO_SITE        9
[PROSITE]       TYR_PHOSPHO_SITE        1
[PROSITE]       PKC_PHOSPHO_SITE        7
[PROSITE]       ASN_GLYCOSYLATION       2
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY     11.20 %

SEQ     MASLGPAAAGEQASGAEAEPGPAGPPPPPSPSSLGPLLPLQREPLYNWQATKASLKERFA
SEG     ....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx...................
PRD     cccccccchhhhhhhhcccccccccccccccccccccccccccccchhhhhhhhhhhhhh

SEQ     FLFNSELLSDVRFVLGKGRGAAAAGGPQRIPAHRFVLAAGSAVFDAMFNGGMATTSAEIE
SEG     ..............xxxxxxxxxxx...................................
PRD     hhhcccccccceeeeeccccccccccccccchhhhheeeccccchhhhhhhhcchhhhhhhee

SEQ     LPDVEPAAFLALLRFLYSDEVQIGPETVMTTLYTAKKYAVPALEAHCVEFLTKHLRADNA
SEG     ............................................................
PRD     eccccchhhhhhhhhhhccceeechhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccch

SEQ     FMLLTQARLFDEPQLASLCLDTIDKSTMDAISAEGFTDIDIDTLCAVLERDTLSIRESRL
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhccccchhhhhh

SEQ     FGAVVRWAEAECQRQQLPVTFGNKQKVLGKALSLIRFPLMTIEEFAAGPAQSGILSDREV
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhcceeeccccccccccccccchhhh

SEQ     VNLFLHFTVNPKPRVEYIDRPRCCLRGKECCINRFQQVESRWGYSGTSDRIRFTVNRRIS
SEG     ............................................................
PRD     hhhhheeeccccceeeeeccccceeecccceeehhhhhhhhcccccccccchhhhhceee

SEQ     IVGFGLYGSIHGPTDYQVNIQIIEYEKKQTLGQNDTGFSCDGTANTFRVMFKEPIEILPN
SEG     ............................................................
PRD     eeeccccccccccchhhhhhhcchhhhhhhhccccccccccccccceeeeeeccceeeccc

SEQ     VCYTACATLKGPDSHYGTKGLKKVVHETPAASKTVFFFFSSPGNNNGTSIEDGQIPEIIF
SEG     ..................................xxxxxx....................
PRD     ccceeeeeccccccccccccceeeeeeecccccceeeeeeeecccccccccccccccceeec

SEQ     YT
SEG     ..
PRD     cc
```

Prosite for DKFZphtes3_35g6.3

PS00001	394->398	ASN_GLYCOSYLATION	PDOC00001
PS00001	466->470	ASN_GLYCOSYLATION	PDOC00001
PS00004	357->361	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	387->391	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	54->57	PKC_PHOSPHO_SITE	PDOC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00005	234->237	PKC_PHOSPHO_SITE	PDOC00005
PS00005	296->299	PKC_PHOSPHO_SITE	PDOC00005
PS00005	348->351	PKC_PHOSPHO_SITE	PDOC00005
PS00005	406->409	PKC_PHOSPHO_SITE	PDOC00005
PS00005	428->431	PKC_PHOSPHO_SITE	PDOC00005
PS00006	14->18	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	115->119	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	217->221	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	281->285	CK2_PHOSPHO_SITE	PDOC00006
PS00006	296->300	CK2_PHOSPHO_SITE	PDOC00006
PS00006	468->472	CK2_PHOSPHO_SITE	PDOC00006
PS00007	430->437	TYR_PHOSPHO_SITE	PDOC00007
PS00008	80->86	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	365->371	MYRISTYL	PDOC00008

PS00008	392->398	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	463->469	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35g6.3)

DKFZphtes3_35k16
----

group: metabolism

DKFZphtes3_35k16 encodes a novel 666 amino acid protein with weak similarity to fatty acid-CoA synthetaseses/ligases.

The novel protein contains a putative AMP-binding domain signature, which is present in enzymes, which act via an ATP-dependent covalent binding of AMP to their substrate. This domain is found in several CoA synthetases, such as acetate-CoA ligase (EC 6.2.1.1), long-chain-fatty-acid-CoA ligase (EC 6.2.1.3),
bile acid-CoA ligase. Therefore it is a new fatty acid-CoA synthetasese/ligase with unknown substrate.

The new protein can find application in modulation of fatty acid metabolism and as a new enzyme for biotechnologic production processes.

similarity to acyl-CoA synthetase

complete cDNA, complete cds, potential start codon at Bp 50,
few EST hits, seems to be a testis specific cDNA,
5 of 6 EST hits are from testis derieved librarys

Sequenced by DKFZ

Locus: unknown

Insert length: 2520 bp
Poly A stretch at pos. 2510, polyadenylation signal at pos. 2490

```
   1 CAGATGTCCC AGCTCCAGTG CTGTGGAGCA TGGTTTCTGC ACACCTGGAA
  51 TGACTGGAAC CCCAAAGACT CAAGAAGGAG CTAAAGATCT TGAAGTAGAC
 101 ATGAATAAAA CAGAAGTTAC TCCCAGGCTG TGGACCACCT GTCGAGATGG
 151 AGAAGTCCTT CTGAGGCTAT CCAAACACGG ACCAGGCCAT GAGACCCCGA
 201 TGACCATCCC TGAATTTTTT CGAGAGTCAG TCAACCGATT TGGAACTTAT
 251 CCAGCCCTCG CATCCAAGAA TGGCAAAAAG TGGGAAATTC TGAATTTCAA
 301 CCAGTACTAT GAGGCTTGTC GGAAGGCTGC AAAATCCTTG ATCAAGCTGG
 351 GTTTGGAGCG TTTCCACGGA GTTGGTATCC TGGGGTTTAA CTCTGCAGAG
 401 TGGTTTATCA CTGCTGTTGG TGCCATCCTA GCCGGGGGTC TTTGTGTTGG
 451 TATTTATGCC ACCAACTCTG CCGAGGCTTG TCAATATGTC ATCACTCATG
 501 CCAAAGTGAA CATCTTGCTG GTTGAGAATG ATCAACAGTT ACAGAAAATC
 551 CTTTCGATTC CACAGAGCAG CCTAGAGCCC CTAAAAGCGA TCATCCAGTA
 601 CAGACTGCCA ATGAAGAAGA ACAACAACTT GTACTCTTGG GATGATTTCA
 651 TGGAACTTGG CAGAAGTATC CCTGACACCC AACTGGAGCA GGTCATCGAG
 701 AGCCAGAAGG CGAATCAATG CGCAGTGCTC ATCTACACTT CAGGGACCAC
 751 AGGCATACCC AAGGGAGTGA TGCTCAGTCA TGACAACATC ACGTGGATTG
 801 CAGGAGCAGT GACAAAGGAC TTTAAACTGA CAGACAAGCA TGAGACGGTG
 851 GTTAGCTACC TCCCACTCAG CCATATTGCA GCACAGATGA TGGACATCTG
 901 GGTACCCATA AAGATTGGGG CGCTCACATA CTTTGCTCAA GCAGATGCTC
 951 TCAAGGGCAC CTTGGTAAGT ACTCTAAAGG AGGTAAAACC TACTGTCTTC
1001 ATTGGAGTGC CTCAAATTTG GGAGAAGATA CATGAGATGG TGAAGAAAAA
1051 TAGTGCCAAG TCCATGGGCT TGAAGAAGAA GGCATTCGTG TGGGCAAGAA
1101 ACATTGGCTT CAAGGTCAAC TCAAAAAAGA TGTTGGGGAA ATATAATACT
1151 CCCGTGAGCT ACCGCATGGC TAAGACTCTC GTGTTCAGCA AAGTCAAGAC
1201 ATCCCTTGGC TTGGATCACT GTCACTCTTT TATCAGTGGG ACTGCGCCCC
1251 TCAACCAAGA GACTGCCGAG TTCTTTCTAA GCTTGGACAT ACCTATAGGC
1301 GAGTTGTATG GGTTGAGTGA GAGCTCGGGA CCCCACACGA TATCCAACCA
1351 GAATAACTAC AGGCTTCTAA GCTGTGGCAA GATCTTGACT GGGTGTAAGA
1401 ATATGCTGTT CCAGCAGAAC AAGGATGGCA TTGGGGAGAT CTGCCTCTGG
1451 GGTAGGCACA TCTTCATGGG CTATCTGGAA AGTGAGACTG AAACTACAGA
1501 GGCCATCGAT GATGAAGGCT GGCTACACTC TGGGGATCTG GGCCAGCTGG
1551 ACGGTCTGGG TTTCCTCTAT GTCACCGGCC ACATCAAAGA AATCCTTATC
1601 ACTGCTGGTG GTGAAAATGT GCCCCCCATT CCTGTTGAGA CCTTGGTTAA
1651 GAAGAAGATC CCCATCATCA GTAACGCCAT GTTAGTAGGA GATAAACTGA
1701 AGTTTCTGAG CATGTTGCTG ACGCTGAAGT GTGAGATGAA TCAGATGAGC
1751 GGAGAACCTC TGGACAAGCT GAACTTCGAG GCCATCAACT TCTGTCGGGG
1801 TCTGGGCAGC CAGGCATCCA CCGTGACTGA GATGGTGAAG CAGCAAGACC
1851 CCCTGGTCTA CAAGGCCATC CAGCAAGGCA TCAATGCTGT GAACCAGGAA
1901 GCCATGAACA ATGCACAGAG GATTGAAAAG TGGGTCATCT TGGAGAAGGA
1951 CTTTTCCATC TATGGTGGAG AGCTAGGTCC AATGATGAAA CTTAAGAGAC
2001 ATTTTGTAGC CCAGAAATAC AAAAAACAAA TTGATCACAT GTACCACTGA
2051 CTGCTTTGAT GGAGCTGCTC TCAGCTGTTC TGATGCCTTC AGCAGGAAGA
2101 CCTCATTGCA ATAAGTGAAA TGCTGCTCTA GGTAGAAGCT CTCCCTGCTG
2151 TTTTTAAGAA GCCACATTCC TCATTGGTCA GTTTCTTGAT TGTTCGTCTG
2201 TTGGAGAGGT GCTCCCTAGA AGAACCTGCC ATACGTTTCA AAGCAATAAA
2251 ATCACTGTAT ATCTTTCTAA GGACCTTCAA GTCATGACTC CAGGGAAGCC
2301 TATTGGGAAG TCTACTAAAA ACTGCCTGAT TTACAAGAAA GACCTGAACT
```

```
2351 TGTGGGCTCC CATTTGATTT TTTTCTCCTC AGGGGACTCA GACATTAGAA
2401 AGAAAAAGCC TCACAGATTT GAAGAACTGG ACCCCCAAAT CAACTCACCT
2451 GCCTGGAAGC AACTGGGAAA CCCTTCCAAT AAGTCCTGAT AATAAAGCAC
2501 TTCAGGGTCC AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 50 bp to 2047 bp; peptide length: 666
Category: similarity to known protein

```
  1 MTGTPKTQEG AKDLEVDMNK TEVTPRLWTT CRDGEVLLRL SKHGPGHETP
 51 MTIPEFFRES VNRFGTYPAL ASKNGKKWEI LNFNQYYEAC RKAAKSLIKL
101 GLERFHGVGI LGFNSAEWFI TAVGAILAGG LCVGIYATNS AEACQYVITH
151 AKVNILLVEN DQQLQKILSI PQSSLEPLKA IIQYRLPMKK NNNLYSWDDF
201 MELGRSIPDT QLEQVIESQK ANQCAVLIYT SGTTGIPKGV MLSHDNITWI
251 AGAVTKDFKL TDKHETVVSY LPLSHIAAQM MDIWVPIKIG ALTYFAQADA
301 LKGTLVSTLK EVKPTVFIGV PQIWEKIHEM VKKNSAKSMG LKKKAFVWAR
351 NIGFKVNSKK MLGKYNTPVS YRMAKTLVFS KVKTSLGLDH CHSFISGTAP
401 LNQETAEFFL SLDIPIGELY GLSESSGPHT ISNQNNYRLL SCGKILTGCK
451 NMLFQQNKDG IGEICLWGRH IFMGYLESET ETTEAIDDEG WLHSGDLGQL
501 DGLGFLYVTG HIKEILITAG GENVPPIPVE TLVKKKIPII SNAMLVGDKL
551 KFLSMLLTLK CEMNQMSGEP LDKLNFEAIN FCRGLGSQAS TVTEMVKQQD
601 PLVYKAIQQG INAVNQEAMN NAQRIEKWVI LEKDFSIYGG ELGPMMKLKR
651 HFVAQKYKKQ IDHMYH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35k16, frame 2

TREMBL:AB014531_1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo sapiens mRNA for KIAA0631 protein, partial cds., N = 1, Score = 1641, P = 8.9e-169

PIR:E70937 probable fadD15 - Mycobacterium tuberculosis (strain H37RV), N = 2, Score = 532, P = 3.6e-62

PIR:H64041 long-chain-fatty-acid--CoA ligase homolog - Haemophilus influenzae (strain Rd KW20), N = 2, Score = 486, P = 6.5e-59

```
>TREMBL:AB014531_1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo
     sapiens mRNA for KIAA0631 protein, partial cds.
          Length = 634

  HSPs:

 Score = 1641 (246.2 bits), Expect = 8.9e-169, P = 8.9e-169
 Identities = 319/628 (50%), Positives = 440/628 (70%)

Query:    38 LRLSKHGPGHETPMTIPEFFRESVNRFGTYPALASKNGKKWEILNFNQYYEACRKAAKSL 97
             LR+    P  + P T+   F E+++++G   AL  K   KWE ++++QYY   R+AAK
Sbjct:     2 LRIDPSCP--QLPYTVHRMFYEALDKYGDLIALGFKRQDKWEHISYSQYYLLARRAAKGF 59

Query:    98 IKLGLERFHGVGILGFNSAEWFITAVGAILAGGLCVGIYATNSAEACQYVITHAKVNILL 157
             +KLGL++ H V ILGFNS EWF +AVG + AGG+  GIY T+S EACQY+       N+++
Sbjct:    60 LKLGLKQAHSVAILGFNSPEWFFSAVGTVFAGGIVTGIYTTSSPEACQYIAYDCCANVIM 119

Query:   158 VENDQQLQKILSIPQSSLEPLKAIIQYRLPM-KKNNNLYSWDDFMELGRSIPDTQLEQVI 216
             V+  +QL+KIL I    L  LKA++ Y+ P   K  N+Y+ ++FMELG  +P+  L+ +I
Sbjct:   120 VDTQKQLEKILKI-WKQLPHLKAVVIYKEPPPNKMANVYTMEEFMELGNEVPEEALDAII 178

Query:   217 ESQKANQCAVLIYTSGTTGIPKGVMLSHDNITWIA--GAVTKDFKLTD-KHETVVSYLPL 273
```

```
                  ++Q+ NQC VL+YTSGTTG PKGVMLS DNITW A  G+    D +   + + E VVSYLPL
Sbjct:   179 DTQQPNQCCVLVYTSGTTGNPKGVMLSQDNITWTARYGSQAGDIRPAEVQQEVVVSYLPL 238

Query:   274 SHIAAQMMDIWVPIKIGALTYFAQADALKGTLVSTLKEVKPTVFIGVPQIWEKIHEMVKK 333
             SHIAAQ+ D+W  I+ GA   FA+ DALKG+LV+TL+EV+PT  +GVP++WEKI E +++
Sbjct:   239 SHIAAQIYDLWTGIQWGAQVCFAEPDALKGSLVNTLREVEPTSHMGVPRVWEKIMERIQE 298

Query:   334 NSAKSMGLKKKAFVWARNIGFKVNSKKMLGKYNTPVSYRMAKTLVFSKVKTSLGLDHCHS 393
              +A+S  +++K  +WA ++  + N     G    P + R+A  LV +KV+ +LG   C
Sbjct:   299 VAAQSGFIRRKMLLWAMSVTLEQNLT-CPGSDLKPFTTRLADYLVLAKVRQALGFAKCQK 357

Query:   394 FISGTAPLNQETAEFFLSLDIPIGELYGLSESSGPHTISNQNNYRLLSCGKILTGCKNML 453
               G AP+  ET  FFL L+I +    YGLSE+SGPH +S+  NYRL S GK++  GC+  L
Sbjct:   358 NFYGAAPMMAETQHFFLGLNIRLYAGYGLSETSGPHFMSSPYNYRLYSSGKLVPGCRVKL 417

Query:   454 FQQNKDGIGEICLWGRHIFMGYLESETETTEAIDDEGWLHSGDLGQLDGLGFLYVTGHIK 513
                Q+ +GIGEICLWGR IFMGYL   E +T EAID+EGWLH+GD G+LD  GFLY+TG +K
Sbjct:   418 VNQDAEGIGEICLWGRTIFMGYLNMEDKTCEAIDEEGWLHTGDAGRLDADGFLYITGRLK 477

Query:   514 EILITAGGENVPPIPVETLVKKKIPIISNAMLVGDKLKFLSMLLTLKCEMNQMSGEPLDK 573
             E++ITAGGENVPP+P+E  VK ++PIISNAML+GD+ KFLSMLLTLKC ++  + +  D
Sbjct:   478 ELIITAGGENVPPVPIEEAVKMELPIISNAMLIGDQRKFLSMLLTLKCTLDPDTSDQTDN 537

Query:   574 LNFEAINFCRGLGSQASTVTEMVKQQDPLVYKAIQQGINAVNQEAMNNAQRIEKWVILEK 633
             L  +A+ FC+ +GS+A+TV+E+++++D  VY+AI++GI  VN  A      I+KW ILE+
Sbjct:   538 LTEQAVEFCQRVGSRATTVSEIIEKKDEAVYQAIEEGIRRVNMNAAARPYHIQKWAILER 597

Query:   634 DFSIYGGELGPMMKLKRHFVAQKYKKQIDHMY 665
             DFSI GGELGP MKLKR  V +KYK  ID  Y
Sbjct:   598 DFSISGGELGPTMKLKRLTVLEKYKGIIDSFY 629
```

Pedant information for DKFZphtes3_35k16, frame 2

Report for DKFZphtes3_35k16.2

[LENGTH] 666
[MW] 74344.97
[pI] 8.67
[HOMOL] TREMBL:AB014531_1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo sapiens mRNA for KIAA0631 protein, partial cds. 1e-176
[FUNCAT] i lipid metabolism [H. influenzae, HI0002] 2e-55
[FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 01.06.13 lipid and fatty-acid transport [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 01.06.07 lipid, fatty-acid and sterol utilization [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YMR246w] 2e-23
[FUNCAT] 06.07 protein modification (glycolsylation, acylation, myristylation, palmitylation, farnesylation and processing) [S. cerevisiae, YMR246w] 2e-23
[BLOCKS] BL00455
[SCOP] d1lci__ 5.19.1.1.1 Luciferase [Firefly (Phontinus pyralis) 1e-49
[EC] 1.13.12.7 Photinus-luciferin 4-monooxygenase (ATP-hydrolysing) 9e-17
[EC] 6.2.1.3 Long-chain-fatty-acid--CoA ligase 4e-34
[EC] 5.1.1.11 Phenylalanine racemase (ATP-hydrolysing) 6e-08
[EC] 6.2.1.12 4-Coumarate--CoA ligase 8e-18
[PIRKW] duplication 6e-07
[PIRKW] phosphopantetheine 3e-12
[PIRKW] multifunctional enzyme 3e-06
[PIRKW] ligase 6e-08
[PIRKW] acid-thiol ligase 4e-34
[PIRKW] transmembrane protein 5e-22
[PIRKW] monooxygenase 9e-17
[PIRKW] hydrolase 4e-34
[PIRKW] peroxisome 9e-15
[PIRKW] antibiotic biosynthesis 3e-12
[PIRKW] isomerase 6e-08
[PIRKW] flavonoid biosynthesis 1e-17
[PIRKW] magnesium 9e-15
[PIRKW] ATP 5e-22
[PIRKW] oxidoreductase 9e-17
[PIRKW] liver 2e-31
[SUPFAM] alpha-aminoadipyl-cysteinyl-valine synthetase 3e-07
[SUPFAM] human long-chain-fatty-acid--CoA ligase 4e-34
[SUPFAM] gramicidin S synthetase I 6e-08
[SUPFAM] peptide synthetase ppsE 7e-06
[SUPFAM] gramicidin S synthetase I repeat homology 3e-12
[SUPFAM] peptide synthetase ppsD 2e-07

```
[SUPFAM]        probable acyl-CoA ligase medium chain 2e-09
[SUPFAM]        acetate--CoA ligase 8e-10
[SUPFAM]        acetate--CoA ligase homology 4e-54
[SUPFAM]        surfactin synthetase 3e-12
[SUPFAM]        4-coumarate--CoA ligase 8e-18
[SUPFAM]        short-chain alcohol dehydrogenase homology 8e-07
[SUPFAM]        acyl carrier protein homology 2e-29
[PROSITE]       MYRISTYL        12
[PROSITE]       AMP_BINDING     1
[PROSITE]       AMIDATION       1
[PROSITE]       CAMP_PHOSPHO_SITE       1
[PROSITE]       CK2_PHOSPHO_SITE        9
[PROSITE]       TYR_PHOSPHO_SITE        3
[PROSITE]       PKC_PHOSPHO_SITE        10
[PROSITE]       ASN_GLYCOSYLATION       2
[PFAM]          AMP-binding enzymes
[KW]            Irregular
[KW]            3D
[KW]            LOW_COMPLEXITY      1.80 %
```

```
SEQ    MTGTPKTQEGAKDLEVDMNKTEVTPRLWTTCRDGEVLLRLSKHGPGHETPMTIPEFFRES
SEG    ............................................................
llci-  ............................................................

SEQ    VNRFGTYPALASKNGKKWEILNFNQYYEACRKAAKSLIKLGLERFHGVGILGFNSAEWFI
SEG    ............................................................
llci-  ............................................................

SEQ    TAVGAILAGGLCVGIYATNSAEACQYVITHAKVNILLVENDQQLQKILSIPQSSLEPLKA
SEG    ............................................................
llci-  ............................................................

SEQ    IIQYRLPMKKNNNLYSWDDFMELGRSIPDTQLEQVIESQKANQCAVLIYTSGTTGIPKGV
SEG    ............................................................
llci-  ............................................................

SEQ    MLSHDNITWIAGAVTKDFKLTDKHETVVSYLPLSHIAAQMMDIWVPIKIGALTYFAQADA
SEG    ............................................................
llci-  ............................................................

SEQ    LKGTLVSTLKEVKPTVFIGVPQIWEKIHEMVKKNSAKSMGLKKKAFVWARNIGFKVNSKK
SEG    ............................................................
llci-  ............................................................

SEQ    MLGKYNTPVSYRMAKTLVFSKVKTSLGLDHCHSFISGTAPLNQETAEFFLSLDIPIGELY
SEG    ............................................................
llci-  ...........................TTTTCEEETTTTCCCHHHHHHHHHHCCCCBCEE

SEQ    GLSESSGPHTISNQNNYRLLSCGKILTGCKNMLFQQNKDGIGEICLWGRHIFMGYLESET
SEG    ............................................................
llci-  ECGGGTTEEEECCCCCCEEEEETTTTEEEEETTTTCEETTEEEEEETTTTCCEETTTHH

SEQ    ETTEAIDDEGWLHSGDLGQLDGLGFLYVTGHIKEILITAGGENVPPIPVETLVKKKIPII
SEG    ..............xxxxxxxxxxxx..................................
llci-  HHHHHBTTTTCEEEEEEEEEETTTTCEEE------ECEEETTEEECHHHHHHHHHT-TTE

SEQ    SNAMLVGDKLKFLSMLLTLKCEMNQMSGEPLDKLNFEAINFCRGLGSQASTVTEMVKQQD
SEG    ............................................................
llci-  EEEEEEE.....................................................

SEQ    PLVYKAIQQGINAVNQEAMNNAQRIEKWVILEKDFSIYGGELGPMMKLKRHFVAQKYKKQ
SEG    ............................................................
llci-  ............................................................

SEQ    IDHMYH
SEG    ......
llci-  ......
```

Prosite for DKFZphtes3_35k16.2

PS00001	19->23	ASN_GLYCOSYLATION	PDOC00001
PS00001	246->250	ASN_GLYCOSYLATION	PDOC00001
PS00004	332->336	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005

PS00005	308->311	PKC_PHOSPHO_SITE	PDOC00005
PS00005	335->338	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	558->561	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	173->177	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	478->482	CK2_PHOSPHO_SITE	PDOC00006
PS00006	591->595	CK2_PHOSPHO_SITE	PDOC00006
PS00007	659->666	TYR_PHOSPHO_SITE	PDOC00007
PS00007	658->666	TYR_PHOSPHO_SITE	PDOC00007
PS00007	597->605	TYR_PHOSPHO_SITE	PDOC00007
PS00008	3->9	MYRISTYL	PDOC00008
PS00008	65->71	MYRISTYL	PDOC00008
PS00008	124->130	MYRISTYL	PDOC00008
PS00008	130->136	MYRISTYL	PDOC00008
PS00008	134->140	MYRISTYL	PDOC00008
PS00008	235->241	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008
PS00008	303->309	MYRISTYL	PDOC00008
PS00008	387->393	MYRISTYL	PDOC00008
PS00008	421->427	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	586->592	MYRISTYL	PDOC00008
PS00009	74->78	AMIDATION	PDOC00009
PS00455	227->239	AMP_BINDING	PDOC00427

Pfam for DKFZphtes3_35k16.2

```
HMM_NAME      AMP-binding enzymes

HMM                *TYRELNERANRLARHLRsekGIrPGDiVgIMMDRSMWMIVaMLGIWKAG
                     + + +E    +A  L+  +G     VGI+  +S +  ++  G + AG
Query          82  NFNQYYEACRKAAKSLI-KLGLERFHGVGILGFNSAEWFITAVGAILAG    129

HMM                GAYVPIDPeYPdERIqYMLEDSGArLLITQrh....HmqRIPdemwwvdH
                   G +V I      +E  QY++ ++ + +L+++ +      +  IP++++  +
Query         130  GLCVGIYATNSAEACQYVITHAKVNILLVENDQQLQKILSIPQSSLEPLK    179

HMM                IiviDWe................WddlWWHedeeNpqpWvdPeDLAYIIY
                   +I++ +                  + ++++ +   E  ++   ++++ A +IY
Query         180  AIIQYRLPMKKNNNLYSWDDFMELGRSIPDTQLEQVIESQKANQCAVLIY    229

HMM                TSGTTGKPKGVMIEHrNIvNycqWMnWRYgMteeDDRILWFtSDpYWFDa
                   TSGTTG PKGVM++H NI+    +   +++ +T+  +++   + +   ++ A
Query         230  TSGTTGIPKGVMLSHDNITWIAGAVTKDFKLTDKHETVVSYLP-LSHIAA    278

HMM                SVWDMFWpLLnGaTLYIpPeEtRrDPerWWqYIqRHgITWWylTPSMFRM
                   +++D++ P+   GA   Y   + ++         +  ++++ ++T+   ++P +++
Query         279  QMMDIWVPIKIGALTYFAQADAL--KGTLVSTLKEVKPTVFIGVPQIWEK    326

HMM                LMpd..............................................
                   + +
Query         327  IHEMVKKNSAKSMGLKKKAFVWARNIGFKVNSKKMLGKYNTPVSYRMAKT    376

HMM                ............psLRhVMFgGEpLsPehWdWWRkrfgfkgRIINMYWPT
                               ++ + +++G  PL++E+++ ++  +    ++I    Y+ +
Query         377  LVFSKVKTSLGLDHCHSFISGTAPLNQETAEFFL-SLD--IPIGELYGLS    423

HMM                ETTVWtTwMrIiPdepeqWrwiPIGRPIpNTqWYIMDdnMQlQPiGViGE
                   E++    T+      +   +    R +++G+ + + + +  +  +N      G IGE
Query         424  ESSGPHTISNQNN--Y---RLLSCGKILTGCKNMLFQQN----KDG-IGE    463

HMM                LYIgGWPGVARGYWNRPELTEERFipNPFWPGEYRrGWNrRMYRTGDLAR
                   +++ G  ++  GY+   + +T E+  + ++                  ++GDL++
Query         464  ICLWG-RHIFMGYLESETETTEAIDDEGW-------------LHSGDLGQ    499

HMM                WlPDGnIEYLGRID.DQVKIRGYRIELGEIEhqLr.qHPgIqEAVV*
                    +  G+++  G I     +    G+++ +  +E+ +  ++P I+ A
Query         500  LDGLGFLYVTGHIKEILITAGGENVPPIPVETLVKKKIPIISNAML     545
```

DKFZphtes3_35k24
----

group: transmembrane protein

DKFZphtes3_35k24 encodes a novel 514 amino acid protein without similarity to known proteins.

The novel protein contains 5 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown ;
membrane regions: 5
Summary DKFZphtes3_35k24 encodes a novel 514 amino acid protein.
No homolouges found in bacteria yeast and C.elegans, specific for mammalians?

unknown

complete cDNA, complete cds, few EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 2706 bp
Poly A stretch at pos. 2696, polyadenylation signal at pos. 2675

```
   1 CCGTGTGCAG TCGCCCCGCG CCCCGCGCGA CCCTTCGGGT AAACTACGAA
  51 CTGGGAGTTC TGAAGAATGG GTAAAGACTT TCGTTACTAT TTCCAGCATC
 101 CCTGGTCTCG CATGATTGTG GCTTACTTGG TGATCTTCTT TAACTTCTTA
 151 ATATTTGCGG AGGACCCAGT TTCTCATAGC CAAACAGAAG CCAATGTTAT
 201 TGTTGTTGGA AACTGTTTTT CATTTGTTAC AAATAAATAC CCTAGAGGAG
 251 TTGGCTGGAG GATTTTGAAG GTGCTTCTAT GGCTACTTGC CATTCTCACA
 301 GGACTAATAG CTGGCAAATT TCTGTTCCAT CAGCGTTTGT TTGGTCAGTT
 351 GCTCCGATTA AAAATGTTTC GAGAAGATCA TGGGTCGTGG ATGACAATGT
 401 TCTTCAGCAC AATTCTCTTT CTCTTCATAT TTTCTCACAT ATACAACACG
 451 ATTCTTCTAA TGGATGGGAA CATGGGAGCA TATATCATTA CAGACTATAT
 501 GGGCATCCGA AATGAAAGTT TCATGAAATT AGCTGCAGTA GGGACCTGGA
 551 TGGGGGACTT TGTCACAGCT TGGATGGTCA CTGATATGAT GCTTCAGGAC
 601 AAACCCTATC CTGACTGGGG AAAATCAGCA AGAGCTTTCT GGAAGAAAGG
 651 AAATGTTAGG ATCACTTTAT TCTGGACAGT TCTTTTTACT CTGACGTCTG
 701 TGGTTGTACT TGTGATTACA ACGGACTGGA TCAGCTGGGA CAAGCTGAAT
 751 CGGGGATTTT TGCCCAGTGA TGAAGTTTCC AGAGCATTCC TTGCTTCTTT
 801 TATCTTGGTC TTTGACCTTC TTATTGTGAT GCAGGACTGG GAATTCCCAC
 851 ATTTCATGGG AGATGTTGAT GTAAATCTCC CTGGTTTGCA CACCCCTCAC
 901 ATGCAGTTCA AGATTCCTTT CTTCCAGAAA ATCTTCAAGG AGGAATATCG
 951 TATTCACATA ACAGGCAAAT GGTTTAACTA TGGAATTATC TTCCTCGTCT
1001 TGATTTTGGA TCTTAATATG TGGAAGAACC AAATATTTTA TAAACCTCAT
1051 GAATATGGGC AATATATCGG CCCGGGGCAG AAGATATATA CAGTGAAAGA
1101 CTCAGAAAGT TTAAAAGATT TGAACAGAAC CAAGCTATCC TGGGAATGGA
1151 GGTCCAATCA CACTAACCCT CGGACTAATA AAACATATGT TGAGGGAGAC
1201 ATGTTCTTAC ACAGCAGGTT CATAGGAGCC AGTCTTGATG TCAAGTGTCT
1251 GGCCTTTGTT CCAAGCCTGA TAGCCTTTGT GTGGTTTGGA TTCTTTATTT
1301 GGTTCTTTGG ACGATTTTTG AAAAATGAGC CACGCATGGA GAATCAAGAC
1351 AAAACTTACA CTCGCATGAA AAGAAAATCT CCATCAGAAC ATAGCAAAGA
1401 CATGGGAATC ACTCGAGAAA ACACCCAGGC TTCAGTAGAA GACCCCTTGA
1451 ATGACCCTTC TTTGGTTTGC ATCAGGTCTG ACTTCAATGA GATCGTCTAC
1501 AAGTCTTCCC ACCTAACCTC GGAAAACTTG AGCTCACAGT TGAACGAATC
1551 TACTAGTGCA ACAGAAGCTG ATCAAGACCC AACGACTTCT AAAAGTACAC
1601 CTACGAACTA GACTCGGAGA TAGACTTGGA GATAACACAA AAAGCAACCT
1651 TGAGTGTAAC TTTAAAAATT TAGTCTTTCC TTTTGTATAT GTAAGGTTTA
1701 CGTAGTGTTA GGTAAAAATA TGAACAATGC CACAACGGTG CTCAACATGC
1751 TTTTTCTAGG ATTCATTGTT TTCTATTTGT ATTATAATAC ACGTGCCTAC
1801 TGTATACTCA ACAGTCCTCT AGAGATTGCT TTTCACAATT GCACAAGCTA
1851 TTACTGACTT TACAGCATAG TGGAAGATTA GCTGATGACC CATGTATCTG
1901 ATGTTCAACC ATAGTGGTGC CTTGAGACAT TAAACTGTTT TTAACTGTAC
1951 CAGAAATGAA GTGTGGAACA GTTACCTAAC CTATTTCACA TGGGCGTTTT
2001 GTATACAACT ATTTTGATCT ACACTTGATG TCTGAGCAGA AAACAGAAAT
2051 AGCTAAATGT GACTCAGGAA GTATCTCTTG GTTTCTTATT CAGCAGCAGA
2101 GTTGGTGACT TTGACAACTG GACTGCAGAG AAACATGGTG ATCACCTTTT
2151 AATTTTTATT GGCTGTCTGC CAAATATAAA TACAGATGCA AAATTCAGTA
2201 ATAGGAGATC CATAACCCAA CATGGGTCAC TACTCGTGAA ATGTGACTTT
2251 CTCCCACCAG TAATTGAAAT GAGGTGATGA TACCTAATTA TGTTTTCCTA
2301 ATTAAAGATA AATTGCTACT TGATTAAAAA TCCTGCCCTT CACCTTTGGG
```

```
2351 AACAAAGGTT AAGAGACACA GTTGGGCGAA CTCTCAAATT TATTGGCATT
2401 TACACAAAGT CCCAGACAAC CAAGGAACTG AAGTTTTCAT CATATGAGAG
2451 CAGCACATCC CACCATTTAC AATATTCGTA TATCTTTCTG CAAATATGGC
2501 TCTGGATAGT GAAAATTGAA AAACATATGC CAACCCTGAG CAAGGGAACT
2551 CCTCAAAAAA TCATGCAGCG GAACCTTGTC AGGTAGAGAA GCCGTGCATG
2601 AAAGAATTTG TTTAATGTCT TGTTTTGCGT ATGTGTTTTT TGTTTTTGTT
2651 TTTTAAGAAC TAAATATTGC ACATTAATAA ATAAGAATTA TACAGCAAAA
2701 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 67 bp to 1608 bp; peptide length: 514
Category: putative protein

```
  1 MGKDFRYYFQ HPWSRMIVAY LVIFFNFLIF AEDPVSHSQT EANVIVVGNC
 51 FSFVTNKYPR GVGWRILKVL LWLLAILTGL IAGKFLFHQR LFGQLLRLKM
101 FREDHGSWMT MFFSTILFLF IFSHIYNTIL LMDGNMGAYI ITDYMGIRNE
151 SFMKLAAVGT WMGDFVTAWM VTDMMLQDKP YPDWGKSARA FWKKGNVRIT
201 LFWTVLFTLT SVVVLVITTD WISWDKLNRG FLPSDEVSRA FLASFILVFD
251 LLIVMQDWEF PHFMGDVDVN LPGLHTPHMQ FKIPFFQKIF KEEYRIHITG
301 KWFNYGIIFL VLILDLNMWK NQIFYKPHEY GQYIGPGQKI YTVKDSESLK
351 DLNRTKLSWE WRSNHTNPRT NKTYVEGDMF LHSRFIGASL DVKCLAFVPS
401 LIAFVWFGFF IWFFGRFLKN EPRMENQDKT YTRMKRKSPS EHSKDMGITR
451 ENTQASVEDP LNDPSLVCIR SDFNEIVYKS SHLTSENLSS QLNESTSATE
501 ADQDPTTSKS TPTN
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35k24, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_35k24, frame 1

Report for DKFZphtes3_35k24.1

```
[LENGTH]      514
[MW]          60185.03
[pI]          8.67
[PROSITE]     MYRISTYL      5
[PROSITE]     CAMP_PHOSPHO_SITE     1
[PROSITE]     CK2_PHOSPHO_SITE      8
[PROSITE]     TYR_PHOSPHO_SITE      1
[PROSITE]     PKC_PHOSPHO_SITE      7
[PROSITE]     ASN_GLYCOSYLATION     6
[KW]          SIGNAL_PEPTIDE 32
[KW]          TRANSMEMBRANE  5
[KW]          LOW_COMPLEXITY    15.37 %

SEQ   MGKDFRYYFQHPWSRMIVAYLVIFFNFLIFAEDPVSHSQTEANVIVVGNCFSFVTNKYPR
SEG   ............................................................
PRD   ccccceeeeeccccchhhhhhhhhhhhhhhhcccccccccccceeeeeeccccceeeeccccc
MEM   ............................................................

SEQ   GVGWRILKVLLWLLAILTGLIAGKFLFHQRLFGQLLRLKMFREDHGSWMTMFFSTILFLF
SEG   ......xxxxxxxxxxxxxxxxx.........................xxxxxxxxxxxx
PRD   cchhhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhccccceeeehhhhhhh
MEM   .....MMMMMMMMMMMMMMMMM...................................MMMMM
```

```
SEQ    IFSHIYNTILLMDGNMGAYIITDYMGIRNESFMKLAAVGTWMGDFVTAWMVTDMMLQDKP
SEG    xxx.........................................................
PRD    hhhhhhhhhhcccccceeeeecccccchhhhhhhhhhccccccccchhhhhhhhhhhccc
MEM    MMMMMMMMMMMM................................................

SEQ    YPDWGKSARAFWKKGNVRITLFWTVLFTLTSVVVLVITTDWISWDKLNRGFLPSDEVSRA
SEG    ..................xxxxxxxxxxxxxxxxxxxxx.....................
PRD    cccccchhhhhhcccceeehhhhhhhhhhhheeeeeccccccccccccccccccchhhhh
MEM    ....................MMMMMMMMMMMMMMMMM......................M

SEQ    FLASFILVFDLLIVMQDWEFPHFMGDVDVNLPGLHTPHMQFKIPFFQKIFKEEYRIHITG
SEG    .......................................xxxxxxxxxxxx.........
PRD    hhhhhhhhhhhhhhhhcccccccccccccccccccccccccchhhhhhhhhhhhhhcccc
MEM    MMMMMMMMMMMMMMMM............................................

SEQ    KWFNYGIIFLVLILDLNMWKNQIFYKPHEYGQYIGPGQKIYTVKDSESLKDLNRTKLSWE
SEG    ............................................................
PRD    ccceeeeeehhhhhhcccccceeecccccccccccccceeeeeccccccccccccchhhh
MEM    ............................................................

SEQ    WRSNHTNPRTNKTYVEGDMFLHSRFIGASLDVKCLAFVPSLIAFVWFGFFIWFFGRFLKN
SEG    ...........................................xxxxxxxxxxxxxx...
PRD    hhcccccccccccccccchhhhhccccccceeeeeehhhheeeeccceeeeeeeeeeccc
MEM    ..................................MMMMMMMMMMMMMMMMM.........

SEQ    EPRMENQDKTYTRMKRKSPSEHSKDMGITRENTQASVEDPLNDPSLVCIRSDFNEIVYKS
SEG    ............................................................
PRD    cccccccccchhhhhcccccccccccceeeccccccccccccccceeeecccccceeeeec
MEM    ............................................................

SEQ    SHLTSENLSSQLNESTSATEADQDPTTSKSTPTN
SEG    ..................................
PRD    cccccccccccccccccccccccccccccccccc
MEM    ..................................
```

Prosite for DKFZphtes3_35k24.1

PS00001	149->153	ASN_GLYCOSYLATION	PDOC00001
PS00001	353->357	ASN_GLYCOSYLATION	PDOC00001
PS00001	364->368	ASN_GLYCOSYLATION	PDOC00001
PS00001	371->375	ASN_GLYCOSYLATION	PDOC00001
PS00001	487->491	ASN_GLYCOSYLATION	PDOC00001
PS00001	493->497	ASN_GLYCOSYLATION	PDOC00001
PS00004	435->439	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	55->58	PKC_PHOSPHO_SITE	PDOC00005
PS00005	187->190	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	348->351	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	507->510	PKC_PHOSPHO_SITE	PDOC00005
PS00006	38->42	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	348->352	CK2_PHOSPHO_SITE	PDOC00006
PS00006	373->377	CK2_PHOSPHO_SITE	PDOC00006
PS00006	438->442	CK2_PHOSPHO_SITE	PDOC00006
PS00006	456->460	CK2_PHOSPHO_SITE	PDOC00006
PS00006	497->501	CK2_PHOSPHO_SITE	PDOC00006
PS00006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS00007	326->334	TYR_PHOSPHO_SITE	PDOC00007
PS00008	48->54	MYRISTYL	PDOC00008
PS00008	79->85	MYRISTYL	PDOC00008
PS00008	106->112	MYRISTYL	PDOC00008
PS00008	134->140	MYRISTYL	PDOC00008
PS00008	159->165	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3_35k24.1)

DKFZphtes3_35n12
----

group: metabolism

DKFZphtes3_35n12 encodes a novel 315 amino acid protein with strong similarity to ADP,ATP carrier T (ANT) proteins.

The novel protein contains three mitochondrial energy transfer signatures and is closely related to the ADP/ATP translocator, or adenine nucleotide translocator (ANT), a protein most abundant in mitochondria. In its functional state, it is a homodimer of 30-kD subunits embedded asymmetrically in the inner mitochondrial membrane. The dimer forms a gated pore through which ADP is moved from the matrix into the cytoplasm.

The new protein can find application in modulation of ADP-transport and energy metabolism in cells/mitochondria.

strong similarity to ADP/ATP carrier proteins

EST hits to mouse and drosophila

Sequenced by DKFZ

Locus: unknown

Insert length: 1803 bp
Poly A stretch at pos. 1793, polyadenylation signal at pos. 1772

```
   1 AGCGTCCCAA GAGCCACTTT CTCGCCAGTA CGATGCTGCA GCGGTTTTCC
  51 GGTTTTCCGC TTCCCTTCAT CGTAGCTCCC GTACTCATTT TTAGCCACTG
 101 CTGCCGGTTT TTATATCCTT CTCCATCATG CATCGTGAGC CTGCGAAAAA
 151 GAAGGCAGAA AAGCGGCTGT TTGACGCCTC ATCCTTCGGG AAGGACCTTC
 201 TGGCCGGCGG AGTCGCGGCA GCTGTGTCCA AGACAGCGGT GGCGCCCATC
 251 GAGCGGGTGA AGCTGCTGCT GCAGGTGCAG GCGTCGTCGA AGCAGATCAG
 301 CCCCGAGGCG CGGTACAAAG GCATGGTGGA CTGCCTGGTG CGGATTCCTC
 351 GCGAGCAGGG TTTCTTCAGT TTTTGGCGTG GCAATTTGGC AAATGTTATT
 401 CGGTATTTTC CAACACAAGC TCTAAACTTT GCTTTTAAGG ACAAATACAA
 451 GCAGCTATTC ATGTCTGGAG TTAATAAAGA AAAACAGTTC TGGAGGTGGT
 501 TTTTGGCAAA CCTGGCTTCT GGTGGAGCTG CTGGGGCAAC ATCCTTATGT
 551 GTAGTATATC CTCTAGATTT TGCCCGAACC CGATTAGGTG TCGATATTGG
 601 AAAAGGTCCT GAGGAGCGAC AATTCAAGGG TTTAGGTGAC TGTATTATGA
 651 AAATAGCAAA ATCAGATGGA ATTGCTGGTT TATACCAAGG GTTTGGTGTT
 701 TCAGTACAGG GCATCATTGT GTACCGAGCC TCTTATTTTG GAGCTTATGA
 751 CACAGTTAAG GGTTTATTAC CAAAGCCAAA GAAAACTCCA TTTCTTGTCT
 801 CCTTTTTCAT TGCTCAAGTT GTGACTACAT GCTCTGGAAT ACTTTCTTAT
 851 CCCTTTGACA CAGTTAGAAG ACGTATGATG ATGCAGAGTG GTGAGGCTAA
 901 ACGGCAATAT AAAGGAACCT TAGACTGCTT TGTGAAGATA TACCAACATG
 951 AAGGAATCAG TTCCTTTTTT CGTGGCGCCT TCTCCAATGT TCTTCGCGGT
1001 ACAGGGGGTG CTTTGGTGTT GGTATTATAT GATAAAATTA AAGAATTCTT
1051 TCATATTGAT ATTGGTGGTA GGTAATCGGG AGAGTAAATT AAGAAATAAC
1101 ATGGATTTAA CTTGTTAAAC ATACAAATTA CATAGCTGCC ATTTGCATAC
1151 ATTTTGATAG TGTTATTGTC TGTATTTTGT TAAAGTGCTA GTTCTGCAAT
1201 AAAGCATACA TTTTTTCAAG AATTTAAATA CTAAAAATCA GATAAATGTG
1251 GATTTTCCTC CCACTTAGAC TCAAACACAT TTTAGTGTGA TATTTCATTT
1301 ATTATAGGTA GTATATTTTA ATTTGTTAGT TTAAAATTCT TTTTATGATT
1351 AAAAATTAAT CATATAATCC TAGATTAATG CTGAAATCTA GGAAATGAAA
1401 GTAGCGTCTT TTAAATTGCT ATTCATTTAA TATACCTGTT TTCCCATCTT
1451 TTGAAGTCAT ATGGTATGAC ATATTTCTTA AAAGCTTATC AATAGATGTC
1501 ATCATATGTG TAGGCAGAAA TAAGCTTTGT TCTATATCTC TTCTAAGACA
1551 GTTGTTATTA CTGTGTATAA TATTTACAGT ATCAGCCTTT GATTATAGAT
1601 GTGATCATTT AAAATTTGAT AATGACTTTA GTGACATTAT AAAACTGAAA
1651 CTGGAAAATA AAATGGCTTA TCTGCTGATG TTTATCTTTA AAATAAATAA
1701 AATCTTGCTA GTGTGAATAT ATCTTAGAAC AAAAGGTATC CTCTTGAAAA
1751 TTAGTTTGTA TATTTTGTTG ACAATAAAGG AAGCTTAACT GTTAAAAAAA
1801 AAA
```

BLAST Results

No BLAST result

Medline entries

96289608:
Molecular biological and quantitative abnormalities of ADP/ATP carrier protein in cardiomyopathic hamsters.

Peptide information for frame 2

```
ORF from 128 bp to 1072 bp; peptide length: 315
Category: strong similarity to known protein
Classification: Metabolism
Prosite motifs: MITOCH_CARRIER (40-50)
MITOCH_CARRIER (145-155)
MITOCH_CARRIER (242-252)


  1 MHREPAKKKA EKRLFDASSF GKDLLAGGVA AAVSKTAVAP IERVKLLLQV
 51 QASSKQISPE ARYKGMVDCL VRIPREQGFF SFWRGNLANV IRYFPTQALN
101 FAFKDKYKQL FMSGVNKEKQ FWRWFLANLA SGGAAGATSL CVVYPLDFAR
151 TRLGVDIGKG PEERQFKGLG DCIMKIAKSD GIAGLYQGFG VSVQGIIVYR
201 ASYFGAYDTV KGLLPKPKKT PFLVSFFIAQ VVTTCSGILS YPFDTVRRRM
251 MMQSGEAKRQ YKGTLDCFVK IYQHEGISSF FRGAFSNVLR GTGGALVLVL
301 YDKIKEFFHI DIGGR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n12, frame 2

PIR:S37210 ADP,ATP carrier protein T1 - mouse, N = 1, Score = 1127, P = 2.7e-114

PIR:A44778 ADP,ATP carrier protein T1 - human, N = 1, Score = 1125, P = 4.4e-114

TREMBL:DMADPATPT_2 product: "ADP/ATP translocase"; Drosophila melanogaster gene encoding ADP/ATP translocase, N = 1, Score = 1124, P = 5.6e-114

PIR:XWBO ADP,ATP carrier protein T1 - bovine, N = 1, Score = 1121, P = 1.2e-113

```
>PIR:S37210 ADP,ATP carrier protein T1 - mouse
         Length = 298

 HSPs:

Score = 1127 (169.1 bits), Expect = 2.7e-114, P = 2.7e-114
Identities = 214/293 (73%), Positives = 248/293 (84%)

Query:    17 ASSFGKDLLAGGVAAAVSKTAVAPIERVKLLLQVQASSKQISPEARYKGMVDCLVRIPRE 76
             A SF KD LAGG+AAAVSKTAVAPIERVKLLLQVQ +SKQIS E +YKG++DC+VRIP+E
Sbjct:     5 ALSFLKDFLAGGIAAAVSKTAVAPIERVKLLLQVQHASKQISAEKQYKGIIDCVVRIPKE 64

Query:    77 QGFFSFWRGNLANVIRYFPTQALNFAFKDKYKQLFMSGVNKEKQFWRWFLANLASGGAAG 136
             QGF SFWRGNLANVIRYFPTQALNFAFKDKYKQ+F+ GV++ KQFWR+F NLASGGAAG
Sbjct:    65 QGFLSFWRGNLANVIRYFPTQALNFAFKDKYKQIFLGGVDRHKQFWRYFAGNLASGGAAG 124

Query:   137 ATSLCVVYPLDFARTRLGVDIGKGPEERQFKGLGDCIMKIAKSDGIAGLYQGFGVSVQGI 196
             ATSLC VYPLDFARTRL  D+GKG  +R+F GLGDC+ KI KSDG+ GLYQGF VSVQGI
Sbjct:   125 ATSLCFVYPLDFARTRLAADVGKGSSQREFNGLGDCLTKIFKSDGLKGLYQGFSVSVQGI 184

Query:   197 IVYRASYFGAYDTVKGLLPKPKKTPFLVSFFIAQVVTTCSGILSYPFDTVRRRMMMQSGE 256
             I+YRA+YFG YDT KG+LP PK    +VS+ IAQ VT  +G++SYPFDTVRRRMMMQSG
Sbjct:   185 IIYRAAYFGVYDTAKGMLPDPKNVHIIVSWMIAQSVTAVAGLVSYPFDTVRRRMMMQSGR 244

Query:   257 --AKRQYKGTLDCFVKIYQHEGISSFFRGAFSNVLRGTGGALVLVLYDKIKEF 307
               A   Y GTLDC+ KI + EG ++FF+GA+SNVLRG GGA VLVLYD+IK++
Sbjct:   245 KGADIMYTGTLDCWRKIAKDEGANAFFKGAWSNVLRGMGGAFVLVLYDEIKKY 297
```

Pedant information for DKFZphtes3_35n12, frame 2

Report for DKFZphtes3_35n12.2

[LENGTH] 315

```
[MW]          35022.03
[pI]          9.91
[HOMOL]       PIR:S37210 ADP,ATP carrier protein T1 - mouse 1e-115
[FUNCAT]      07.16 purine and pyrimidine transporters    [S. cerevisiae, YBL030c] 2e-72
[FUNCAT]      08.04 mitochondrial transport        [S. cerevisiae, YBL030c] 2e-72
[FUNCAT]      30.16 mitochondrial organization     [S. cerevisiae, YBL030c] 2e-72
[FUNCAT]      01.03.19 nucleotide transport        [S. cerevisiae, YBL030c] 2e-72
[FUNCAT]      01.07.10 transport of vitamins, cofactors, and prosthetic groups    [S.
cerevisiae, YIL006w] 2e-14
[FUNCAT]      07.99 other transport facilitators   [S. cerevisiae, YIL006w] 2e-14
[FUNCAT]      01.05.07 carbohydrate transport      [S. cerevisiae, YPR021c] 5e-14
[FUNCAT]      07.07 sugar and carbohydrate transporters    [S. cerevisiae, YPR021c] 5e-14
[FUNCAT]      07.04.07 anion transporters (cl, so4, po4, etc.)    [S. cerevisiae, YKL120w]
1e-13
[FUNCAT]      02.13 respiration     [S. cerevisiae, YBR192w] 4e-13
[FUNCAT]      01.05.04 regulation of carbohydrate utilization     [S. cerevisiae, YJR095w]
6e-12
[FUNCAT]      13.04 homeostasis of other ions      [S. cerevisiae, YLR348c] 4e-10
[FUNCAT]      01.04.07 phosphate transport  [S. cerevisiae, YLR348c] 4e-10
[FUNCAT]      01.01.07 amino-acid transport        [S. cerevisiae, YOR130c] 1e-06
[FUNCAT]      07.10 amino-acid transporters        [S. cerevisiae, YOR130c] 1e-06
[FUNCAT]      99 unclassified proteins     [S. cerevisiae, YPR128c] 2e-06
[FUNCAT]      04.05.03 mrna processing (splicing)   [S. cerevisiae, YKR052c] 2e-06
[BLOCKS]      BL00215B Mitochondrial energy transfer proteins
[BLOCKS]      BL00215A Mitochondrial energy transfer proteins
[PIRKW]       duplication 1e-115
[PIRKW]       phosphate transport 2e-09
[PIRKW]       heart 3e-24
[PIRKW]       transmembrane protein 1e-115
[PIRKW]       mitochondrial inner membrane 7e-72
[PIRKW]       transport protein 4e-08
[PIRKW]       acetylated amino end 1e-115
[PIRKW]       adipose tissue 5e-13
[PIRKW]       mitochondrion 1e-115
[PIRKW]       alternative splicing 2e-09
[PIRKW]       methylated amino acid 1e-115
[PIRKW]       chloroplast 2e-14
[PIRKW]       homodimer 1e-115
[SUPFAM]      hypothetical protein YFR045w 3e-07
[SUPFAM]      ADP,ATP carrier protein 1e-115
[SUPFAM]      Bt1 protein 2e-14
[SUPFAM]      ADP,ATP carrier protein repeat homology 1e-115
[SUPFAM]      probable carrier protein YPR021c 1e-12
[PROSITE]     MITOCH_CARRIER 3
[PFAM]        Mitochondrial carrier proteins
[KW]          TRANSMEMBRANE 2
[KW]          LOW_COMPLEXITY     4.76 %
```

```
SEQ   MHREPAKKKAEKRLFDASSFGKDLLAGGVAAAVSKTAVAPIERVKLLLQVQASSKQISPE
SEG   ............................................................
PRD   ccchhhhhhhhhhhhhchhhhhhhhhchhhhhhhhhhhcchhhhhhhhhhhhhhhhhhh
MEM   ............................................................

SEQ   ARYKGMVDCLVRIPREQGFFSFWRGNLANVIRYFPTQALNFAFKDKYKQLFMSGVNKEKQ
SEG   ............................................................
PRD   hhhhhhhheeeeccccceeeeecccccceeeeeccccchhhhhhhhhhhhhhcccccccc
MEM   ............................................................

SEQ   FWRWFLANLASGGAAGATSLCVVYPLDFARTRLGVDIGKGPEERQFKGLGDCIMKIAKSD
SEG   .....xxxxxxxxxxxxxxx........................................
PRD   eeeecccccccccccccceeeeeeecccchhhhhhhhhhcccccchhhhhhcccceeeeeccc
MEM   ............................................................

SEQ   GIAGLYQGFGVSVQGIIVYRASYFGAYDTVKGLLPKPKKTPFLVSFFIAQVVTTCSGILS
SEG   ............................................................
PRD   ccccccccceeecccceeehhhhhccccccccccccccccccchhhhhhhhhhheeeeec
MEM   ....MMMMMMMMMMMMMMMMMMMMMMM.................MMMMMMMMMMMMMMMM

SEQ   YPFDTVRRRMMQSGEAKRQYKGTLDCFVKIYQHEGISSFFRGAFSNVLRGTGGALVLVL
SEG   ............................................................
PRD   ccchhhhhhhhhcccceeeeccccchhhhhhhhhcccccccccchhhhhcccccceeeeee
MEM   MMMMMMMMMM..................................................

SEQ   YDKIKEFFHIDIGGR
SEG   ...............
PRD   hhhhhhheeeecccc
MEM   ...............
```

Prosite for DKFZphtes3_35n12.2

PS00215	40->50	MITOCH_CARRIER	PDOC00189
PS00215	145->155	MITOCH_CARRIER	PDOC00189
PS00215	242->252	MITOCH_CARRIER	PDOC00189

Pfam for DKFZphtes3_35n12.2

```
HMM_NAME    Mitochondrial carrier proteins

HMM              *pFwkdFLAGGIAGmMeHTvMFPIDtIKTRMQlQgEMpM..ahpRYkGMI
                  +F+KD+LAGG+A+++++T+++PI+++K+++Q+Q  +++      RYKGM+
Query       19    SFGKDLLAGGVAAAVSKTAVAPIERVKLLLQVQASSKQISPEARYKGMV    67

HMM              dCFRwIwkNEGWRGLWRGLgANvIRYIPqWaIRFGFYEFMKeMFiDyfge
                 DC+ +I++++G++++WRG++ANVIRY+P++A++F+F++ +K +F + +++
Query       68   DCLVRIPREQGFFSFWRGNLANVIRYFPTQALNFAFKDKYKQLFMSGVNK    117

HMM              ddnyWmWFwmnYMaGsmAGEwisvIitYPMWvVKTRLQaDqkHphsQp.R
                   ++W+WF+ N+++G++AG ++S+ ++YP+++++TRL  D  +++++  R
Query       118  EKQFWRWFLANLASGGAAG-ATSLCVVYPLDFARTRLGVD--IGKGPEER    164

HMM              hYNGvWNcWrkIYReEGgFkGLYRGWtPTWMRMIPYqmiYFfvYEtLKeW
                 +++G+ +C  KI +++G ++GLY+G++ +++++I+Y++ YF++Y+T K +
Query       165  QFKGLGDCIMKIAKSDG-IAGLYQGFGVSVQGIIVYRASYFGAYDTVKGL    213

HMM              lynYtgYnPgprelCMddsPwWhWiIgWmIAGMiaWivSYPfDVVRTRMM
                 L          +++   +  ++++++I++ ++ ++++I+SYPFD+VR+RMM
Query       214  LP---------KPK--KTPFLVSFFIAQVVT-TCSGILSYPFDTVRRRMM    251

HMM              Mdsm.edhkYqSmlDCWMqIYKnEGFkGFWKGFWPRIMRiMPWtAIMFmI
                 M+S+ ++++Y+++LDC+++IY++EG+ +F++G+   +++R+  ++A+++++
Query       252  MQSGEAKRQYKGTLDCFVKIYQHEGISSFFRGAFSNVLRGT-GGALVLVL    300

HMM              YEqMKwFL*
                 Y+ +K+F+
Query       301  YDKIKEFF    308
```

DKFZphtes3_35n24
----

group: testes derived

DKFZphtes3_35n24 encodes a novel 365 amino acid protein without similarity to known proteins.

The novel protein contains a Prosite Ig(Immunoglubulin)-MHC pattern. This pattern represents a domain, approximately one hundred amino acids long and including a conserved intra-domain disulfide bond (ÿIg domainï). Thus, the novel protein is a new member of the Ig-superfamily. No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1589 bp
Poly A stretch at pos. 1579, polyadenylation signal at pos. 1560

```
   1 CGATCGTCAC GTGACGCCGG GGTTCAGCGT ATCCTTGCTG GGCAACCGTC
  51 TTAGAGACCA GCACTGCTGG CTGCACCATG AATGTGATCT ACCCACTGGC
 101 AGTCCCCAAG GGGCGCAGAC TCTGCTGTGA GGTGTGCGAA GCCCCAGCCG
 151 AGCGGGTGTG CGCGGCCTGC ACAGTCACTT ATTACTGTGG GGTGGTACAT
 201 CAGAAGGCTG ACTGGGACAG CATCCATGAG AAAATATGTC AGCTCTTGAT
 251 TCCACTGCGC ACTTCCATGC CCTTCTACAA TTCAGAGGAA GAACGGCAGC
 301 ATGGCCTGCA GCAGCTGCAG CAGCGGCAGA AGTATTTGAT TGAATTCTGC
 351 TACACCATAG CCCAGAAATA CCTCTTTGAA GGGAAACACG AAGATGCTGT
 401 ACCAGCAGCT TTGCAGTCCC TTCGCTTCCG TGTGAAGCTG TATGGCCTGA
 451 GCTCCGTAGA GCTTGTGCCT GCTTACCCGC TGTTGGCCGA GGCCAGCCTT
 501 GGTCTGGGCC GAATCGTTCA GGCTGAAGAA TATCTATTCC AAGCCCAGTG
 551 GACAGTCCTC AAATCAACTG ACTGTAGTAA TGCCACCCAC TCTTTACTGC
 601 ATCGGAATCT GGGACTTCTC TATATAGCTA AGAAAAACTA TGAAGAGGCC
 651 CGTTATCATC TGGCCAATGA TATTTATTTT GCCAGTTGTG CATTTGGAAC
 701 AGAGGACATT AGGACTTCAG GAGGCTACTT CCACCTGGCT AATATATTCT
 751 ATGACCTTAA AAAGTTGGAC CTGGCAGACA CATTGTACAC CAAGGTCTCT
 801 GAGATCTGGC ATGCATATTT GAACAATCAC TATCAAGTCC TCTCACAGGC
 851 TCACATCCAA CAAATGGATT TACTGGGCAA ACTATTTGAG AATGACACTG
 901 GCTTGGATGA AGCCCAAGAA GCAGAAGCCA TTCGCATCCT GACTTCAATC
 951 TTGAACATTC GAGAATCTAC ATCTGACAAA GCCCCCCAAA AAACCATCTT
1001 TGTTCTGAAG ATCCTGGTCA TGCTTTACTA CCTGATGATG AATTCTTCAA
1051 AGGCACAGGA ATATGGCATG AGGGCCCTCA GTCTAGCCAA AGAACAACAG
1101 CTTGATGTCC ATGAGCAAAG CACCATTCAA GAGTTATTAA GTCTCATTTC
1151 AACTGAAGAC CATCCCATTA CTTAGTGACC CATGAGCTCT GCATCAAGGG
1201 TTATTCCAGG GGCTACTGAA GATCTAATAT ATTCCAGCCT TGCACAACTG
1251 CTTTGAGGTA CTGTAGACTG CTGAAGTTTC CACCCTCTTC CCCTGGGATT
1301 GCACACATAG CTGTTATTTT TTTCTTACAC AGCATATTAA GGGAATATAA
1351 AGCTTTAGGC ATAGAAATCA CTAAAAACTG TGTTTGTCAT GACCTTTGTA
1401 CTTGATTTAT CATGACTTTG TATGACTGAG TAATATGTAG TCAGATCACT
1451 AATATGGTAT TTGTAATTAA ACTACAAATA GTTTGTCATT TCCCAGAAGT
1501 CTTCCAACGA TGCATGTTTC ATACACTTTT GCTAAAGGAG GGGTAAAGGA
1551 GGGGGTAGGG AATAAAGCTA TATTGGAACA AAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 78 bp to 1172 bp; peptide length: 365
Category: putative protein

Prosite motifs: IG_MHC (35-42)

```
  1 MNVIYPLAVP KGRRLCCEVC EAPAERVCAA CTVTYYCGVV HQKADWDSIH
 51 EKICQLLIPL RTSMPFYNSE EERQHGLQQL QQRQKYLIEF CYTIAQKYLF
101 EGKHEDAVPA ALQSLRFRVK LYGLSSVELV PAYPLLAEAS LGLGRIVQAE
151 EYLFQAQWTV LKSTDCSNAT HSLLHRNLGL LYIAKKNYEE ARYHLANDIY
201 FASCAFGTED IRTSGGYFHL ANIFYDLKKL DLADTLYTKV SEIWHAYLNN
251 HYQVLSQAHI QQMDLLGKLF ENDTGLDEAQ EAEAIRILTS ILNIRESTSD
301 KAPQKTIFVL KILVMLYYLM MNSSKAQEYG MRALSLAKEQ QLDVHEQSTI
351 QELLSLISTE DHPIT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n24, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_35n24, frame 3

Report for DKFZphtes3_35n24.3

```
[LENGTH]        365
[MW]            41768.24
[pI]            5.82
[BLOCKS]        BL00273 Heat-stable enterotoxins proteins
[PROSITE]       MYRISTYL        1
[PROSITE]       IG_MHC  1
[PROSITE]       AMIDATION       1
[PROSITE]       CK2_PHOSPHO_SITE        7
[PROSITE]       TYR_PHOSPHO_SITE        4
[PROSITE]       PKC_PHOSPHO_SITE        3
[PROSITE]       ASN_GLYCOSYLATION       3
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY      4.11 %
```

```
SEQ     MNVIYPLAVPKGRRLCCEVCEAPAERVCAACTVTYYCGVVHQKADWDSIHEKICQLLIPL
SEG     ............................................................
PRD     ccceeeeecccccceeeeeeeehhhhhhhheeeeeeeeeecccccccchhhhhhhhheeec

SEQ     RTSMPFYNSEEERQHGLQQLQQRQKYLIEFCYTIAQKYLFEGKHEDAVPAALQSLRFRVK
SEG     .........xxxxxxxxxxxxxxx....................................
PRD     cccccccchhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhcccccccchhhhhhhhhhh

SEQ     LYGLSSVELVPAYPLLAEASLGLGRIVQAEEYLFQAQWTVLKSTDCSNATHSLLHRNLGL
SEG     ............................................................
PRD     hhccceeeeccccchhhhhcccccchhhhhhhhhhhhhhhcccccccccccccccccccc

SEQ     LYIAKKNYEEARYHLANDIYFASCAFGTEDIRTSGGYFHLANIFYDLKKLDLADTLYTKV
SEG     ............................................................
PRD     eeeehhhhhhhhhhhhhheeeeecccccccccccccceeehhhhhhhhhhhccceeeeeh

SEQ     SEIWHAYLNNHYQVLSQAHIQQMDLLGKLFENDTGLDEAQEAEAIRILTSILNIRESTSD
SEG     ............................................................
PRD     hhhhhhhhccchhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhccccc

SEQ     KAPQKTIFVLKILVMLYYLMMNSSKAQEYGMRALSLAKEQQLDVHEQSTIQELLSLISTE
SEG     ............................................................
PRD     cccccceeehhhhhhhhhhhhccccchhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhcc

SEQ     DHPIT
SEG     .....
PRD     ccccc
```

Prosite for DKFZphtes3_35n24.3

PS00001	168->172	ASN_GLYCOSYLATION	PDOC00001
PS00001	272->276	ASN_GLYCOSYLATION	PDOC00001
PS00001	322->326	ASN_GLYCOSYLATION	PDOC00001
PS00005	114->117	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005

PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	69->73	CK2_PHOSPHO_SITE	PDOC00006
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	274->278	CK2_PHOSPHO_SITE	PDOC00006
PS00006	297->301	CK2_PHOSPHO_SITE	PDOC00006
PS00006	349->353	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00007	85->93	TYR_PHOSPHO_SITE	PDOC00007
PS00007	186->194	TYR_PHOSPHO_SITE	PDOC00007
PS00007	186->194	TYR_PHOSPHO_SITE	PDOC00007
PS00007	185->194	TYR_PHOSPHO_SITE	PDOC00007
PS00008	275->281	MYRISTYL	PDOC00008
PS00009	11->15	AMIDATION	PDOC00009
PS00290	35->42	IG_MHC	PDOC00262

(No Pfam data available for DKFZphtes3_35n24.3)

DKFZphtes3_35n9
----

group: metabolism

DKFZphftes3_35n9 encodes a novel 607 amino acid protein which is a splice variant of human carboxylesterase (EC 3.1.1.1).

The novel protein contains both, one carboxylesterase B1 and one B2 pattern. In comparison to EC 3.1.1.1, DKFZphtes3_35n9 shows a N-terminal extension and aa 458-474 are missing.

The new protein can find application in modulation of carboxylester metabolism and as a new enzyme for biotechnologic production processes.

carboxylesterase, splice variant

5' extension of mRNA and N-terminal elongation of protein (64 aa),
missing exon! aa 458-474 of JC5408 are missing

Sequenced by DKFZ

Locus: unknown

Insert length: 2888 bp
Poly A stretch at pos. 2878, no polyadenylation signal found

```
   1 CTCGGCCTGA GGTGCGAGAG AAGCGGTGAC CGCGGCCCTG GCTGCTCGGA
  51 CCCGGGAACA TGATGGTCGC TGGAGCAGAA GGCGCTGAGA AGGGACCACG
 101 GCGGCGCTGG GTCGTGCGAG CCAGTAGCGG GCTGAAACGT AGAGGCCAGA
 151 ACCAGGTCTC AGGGGGCACT AAAGGCGGTC GGAGGTAATC CCCACACCGC
 201 TTCCTCCTGG AAGTCAGGCT GGCCGGGAGC TCCCGTATCC AGGACGGTTG
 251 GTCGCCTCTG GCCTGGCAGG GATCCTAGTG TCTCGGGACC TCCCGGTGAC
 301 GCGCCTGCCT CCCCTGCTGC ACCATAGGCC CGGGAGTACG GCGTCCCCAC
 351 AGCTTGGACC GGCAGGGGCT CGTGAAATGT TTGTCAAGTG GATAAATGAC
 401 CATGGCCGTG GTCTCCGCGG GAGGTGAGGA AACTGAAAGC CACCGAGGAA
 451 AAGGGGGGCG CTCCTTAAGA AGTGCCGCGG TCACGTGTAC GTTTCAAAAG
 501 AATGGCGTGA CTGAGTAGGG AGGGGACCGC GGAGACCCTC AGACCCTGGA
 551 CTGTAAGGAG ATGAGGGGCC GTGAAGGGGA ACCCAGGAAA CTGAGTCCTG
 601 AAAGCAAGGA GGAACTTCCA GAATGAAGGG CGCCGACACT CCTTCCTGCC
 651 TTTGCTCAAG CGGTTCCTTC ACCCCGATCA AGTTCCTTCC CATTTCTCCA
 701 TCTGGGGGAT CCTGAACGTG CACATCCTCA GAGAAGCCCT CCTGGGGTCT
 751 CCAATTCTAG TTTATTGCCC CCTCCTATCG ATCCCCCAGC GCGCTCATCG
 801 GGCCTGTGGA CAAGGACAGG TTTGAAGAGA GGATTCCCTG GATCGCGGAA
 851 GGGCTGCAGG AATGGCACAG CCCCTTCCGA GGATGCCAAA GGAGCCCGGG
 901 CAAAGGAAAG TGGCCGTGCC CGGGCCTGCC TACCACTAGA TCCCCACCCA
 951 CCTATGACTG CTCAGTCCCG CTCTCCTACC ACACCCACCT TTCCCGGCCC
1001 AAGCCAGCGC ACCCCGCTGA CTCCCTGCCC AGTCCAAACT CCAAGGCTGG
1051 GCAAGGCACT GATCCACTGC TGGACAGACC CGGGGCAGCC TCTGGGTGAA
1101 CAGCAGCGTG TCCGCCGGCA GCGAACCGAG ACCAGCGAGC CGACCATGCG
1151 GCTGCACAGA CTTCGTGCGC GGCTGAGCGC GGTGGCCTGT GGGCTTCTGC
1201 TGCTTCTTGT CCGGGGCCAG GGCCAGGACT CAGCCAGTCC CATCCGGACC
1251 ACACACACGG GGCAGGTGCT GGGGAGTCTT GTCCATGTGA AGGGCGCCAA
1301 TGCCGGGGTC CAAACCTTCC TGGGAATTCC ATTTGCCAAG CCACCTCTAG
1351 GTCCGCTGCG ATTTGCACCC CCTGAGCCCC CTGAATCTTG GAGTGGTGTG
1401 AGGGATGGAA CCACCCATCC GGCCATGTGT CTACAGGACC TCACCGCAGT
1451 GGAGTCAGAG TTTCTTAGCC AGTTCAACAT GACCTTCCCT TCCGACTCCA
1501 TGTCTGAGGA CTGCCTGTAC CTCAGCATCT ACACGCCGGC CCATAGCCAT
1551 GAAGGCTCTA ACCTGCCGGT GATGGTGTGG ATCCACGGTG GTGCGCTTGT
1601 TTTTGGCATG GCTTCCTTGT ATGATGGTTC CATGCTGGCT GCCTTGGAGA
1651 ACGTGGTGGT GGTCATCATC CAGTACCGCC TGGGTGTCCT GGGCTTCTTC
1701 AGCACTGGAG ACAAGCACGC AACCGGCAAC TGGGGCTACC TGGACCAAGT
1751 GGCTGCACTA CGCTGGGTCC AGCAGAATAT CGCCCACTTT GGAGGCAACC
1801 CTGACCGTGT CACCATTTTT GGCGAGTCTG CGGGTGGCAC GAGTGTGTCT
1851 TCGCTTGTTG TGTCCCCCAT ATCCCAAGGA CTCTTCCACG GAGCCATCAT
1901 GGAGAGTGGC GTGGCCCTCC TGCCCGGCCT CATTGCCAGC TCAGCTGATG
1951 TCATCTCCAC GGTGGTGGCC AACCTGTCTG CCTGTGACCA AGTTGACTCT
2001 GAGGCCCTGG TGGGCTGCCT GCGGGGCAAG AGTAAAGAGG AGATTCTTGC
2051 AATTAACAAG CCTTTCAAGA TGATCCCCGG AGTGGTGGAT GGGGTCTTCC
2101 TGCCCAGGCA CCCCCAGGAG CTGCTGGCCT CTGCCGACTT TCAGCCTGTC
2151 CCTAGCATTG TTGGTGTCAA CAACAATGAA TTCGGCTGGC TCATCCCCAA
2201 GGTCATGAGG ATCTATGATA CCCAGAAGGA AATGGACAGA GAGGCCTCCC
2251 AGGCTGCTCT GCAGAAAATG TTAACGCTGC TGATGTTGCC TCCTACATTT
2301 GGTGACCTGC TGAGGGAGGA GTACATTGGG GACAATGGGG ATCCCCAGAC
2351 CCTCCAAGCG CAGTTCCAGG AGATGATGGC GGACTCCATG TTTGTGATCC
2401 CTGCACTCCA AGTAGCACAT TTTCAGTGTT CCCGGGCCCC TGTGTACTTC
2451 TACGAGTTCC AGCATCAGCC CAGCTGGCTC AAGAACATCA GGCCACCGCA
2501 CATGAAGGCA GACCATGTTA AATTCACTGA GGAAGAGGAG CAGCTAAGCA
2551 GGAAGATGAT GAAGTACTGG GCCAACTTTG CGAGAAATGG GAACCCCAAT
2601 GGCGAGGGTC TGCCACACTG GCCGCTGTTC GACCAGGAGG AGCAATACCT
```

```
2651 GCAGCTGAAC CTACAGCCTG CGGTGGGCCG GGCTCTGAAG GCCCACAGGC
2701 TCCAGTTCTG GAAGAAGGCG CTGCCCCAAA AGATCCAGGA GCTCGAGGAG
2751 CCTGAAGAGA GACACACAGA GCTGTAGCTC CCTGTGCCGG GGAGGAGGGG
2801 GTGGGTTCGC TGACAGGCGA GGGTCAGCCT GCTGTGCCCA CACACACCCA
2851 CTAAGGAGAA AGAAGTTGAT TCCTTCATAA AAAAAAAA
```

BLAST Results

Entry D50579 from database EMBL:
Homo sapiens mRNA for carboxylesterase, complete cds.
Score = 7197, P = 0.0e+00, identities = 1441/1443

Entry JC5408 from database PIR:
carboxylesterase (EC 3.1.1.1) - human
Score = 2808, P = 1.2e-291, identities = 542/559, positives = 543/559, frame +3

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 954 bp to 2774 bp; peptide length: 607
Category: known protein
Classification: Metabolism
Prosite motifs: CARBOXYLESTERASE_B_1 (279-295)
CARBOXYLESTERASE_B_2 (185-196)

```
  1 MTAQSRSPTT PTFPGPSQRT PLTPCPVQTP RLGKALIHCW TDPGQPLGEQ
 51 QRVRRQRTET SEPTMRLHRL RARLSAVACG LLLLLVRGQG QDSASPIRTT
101 HTGQVLGSLV HVKGANAGVQ TFLGIPFAKP PLGPLRFAPP EPPESWSGVR
151 DGTTHPAMCL QDLTAVESEF LSQFNMTFPS DSMSEDCLYL SIYTPAHSHE
201 GSNLPVMVWI HGGALVFGMA SLYDGSMLAA LENVVVVIIQ YRLGVLGFFS
251 TGDKHATGNW GYLDQVAALR WVQQNIAHFG GNPDRVTIFG ESAGGTSVSS
301 LVVSPISQGL FHGAIMESGV ALLPGLIASS ADVISTVVAN LSACDQVDSE
351 ALVGCLRGKS KEEILAINKP FKMIPGVVDG VFLPRHPQEL LASADFQPVP
401 SIVGVNNNEF GWLIPKVMRI YDTQKEMDRE ASQAALQKML TLLMLPPTFG
451 DLLREEYIGD NGDPQTLQAQ FQEMMADSMF VIPALQVAHF QCSRAPVYFY
501 EFQHQPSWLK NIRPPHMKAD HVKFTEEEEQ LSRKMMKYWA NFARNGNPNG
551 EGLPHWPLFD QEEQYLQLNL QPAVGRALKA HRLQFWKKAL PQKIQELEEP
601 EERHTEL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35n9, frame 3

PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human, N = 1, Score = 2808, P = 1.9e-292

TREMBL:HSU60553_1 gene: "hCE-2"; product: "carboxylesterase"; Human carboxylesterase (hCE-2) mRNA, complete cds., N = 1, Score = 2761, P = 1.8e-287

PIR:A34329 60K esterase (EC 3.1.1.-) isoform 2 - rabbit, N = 1, Score = 1985, P = 3.1e-205

TREMBL:D50580_1 product: "carboxylesterase precursor"; Rattus norvegicus mRNA for carboxylesterase, partial cds., N = 1, Score = 1984, P = 4e-205

>PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human
Length = 559

HSPs:

Score = 2808 (421.3 bits), Expect = 1.9e-292, P = 1.9e-292

```
Identities = 542/559 (96%), Positives = 543/559 (97%)

Query:   65 MRLHRLRARLSAVACGLLLLLVRGQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQTFLG 124
            MRLHRLRARLSAVACGLLLLLVRGQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQTFLG
Sbjct:    1 MRLHRLRARLSAVACGLLLLLVRGQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQTFLG 60

Query:  125 IPFAKPPLGPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTFPSDSMS 184
            IPFAKPPLGPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTFPSDSMS
Sbjct:   61 IPFAKPPLGPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTFPSDSMS 120

Query:  185 EDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVVIIQYRLG 244
            EDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVVIIQYRLG
Sbjct:  121 EDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVVIIQYRLG 180

Query:  245 VLGFFSTGDKHATGNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSLVVS 304
            VLGFFSTGDKHATGNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSLVVS
Sbjct:  181 VLGFFSTGDKHATGNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSLVVS 240

Query:  305 PISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCLRGKSKEEI 364
            PISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCLRGKSKEEI
Sbjct:  241 PISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCLRGKSKEEI 300

Query:  365 LAINKPFKMIPGVVDGVFLPRHPQELLASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQ 424
            LAINKPFKMIPGVVDGVFLPRHPQELLASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQ
Sbjct:  301 LAINKPFKMIPGVVDGVFLPRHPQELLASADFQPVPSIVGVNNNEFGWLIPKVMRIYDTQ 360

Query:  425 KEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMFVIPA 484
            KEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMFVIPA
Sbjct:  361 KEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMFVIPA 420

Query:  485 LQVAHFQCSRAPVYFYEFQHQPSWLKNIRPPHMKADH----------------VKFTEEE 528
            LQVAHFQCSRAPVYFYEFQHQPSWLKNIRPPHMKADH                +KFTEEE
Sbjct:  421 LQVAHFQCSRAPVYFYEFQHQPSWLKNIRPPHMKADHGDELPFVFRSFFGGNYIKFTEEE 480

Query:  529 EQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQYLQLNLQPAVGRALKAHRLQFWKK 588
            EQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQYLQLNLQPAVGRALKAHRLQFWKK
Sbjct:  481 EQLSRKMMKYWANFARNGNPNGEGLPHWPLFDQEEQYLQLNLQPAVGRALKAHRLQFWKK 540

Query:  589 ALPQKIQELEEPEERHTEL 607
            ALPQKIQELEEPEERHTEL
Sbjct:  541 ALPQKIQELEEPEERHTEL 559
```

Pedant information for DKFZphtes3_35n9, frame 3

Report for DKFZphtes3_35n9.3

```
[LENGTH]    607
[MW]        67051.20
[pI]        6.11
[HOMOL]     PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human 0.0
[BLOCKS]    BL01173A Lipolytic enzymes "G-D-X-G" family,  histidine
[BLOCKS]    BL00122G
[BLOCKS]    BL00122F
[BLOCKS]    BL00122E
[BLOCKS]    BL00122D Carboxylesterases type-B serine proteins
[BLOCKS]    BL00122C Carboxylesterases type-B serine proteins
[BLOCKS]    BL00122B Carboxylesterases type-B serine proteins
[BLOCKS]    BL00122A Carboxylesterases type-B serine proteins
[SCOP]      d1akn__ 3.56.1.1.4 Bile-salt activated lipase  [Bovine (Bos taurus 1e-158
[SCOP]      d2ack__ 3.56.1.1.1 Acetylcholinesterase  [Electric ray (Torped 1e-170
[SCOP]      d1thg__ 3.56.1.9.7 type-B carboxylesterase/lipase  [fungu 1e-149
[EC]        3.1.1.13 Sterol esterase 1e-52
[EC]        3.1.1.7 Acetylcholinesterase 5e-74
[EC]        3.1.1.1 Carboxylesterase 0.0
[EC]        3.1.1.8 Cholinesterase 5e-68
[EC]        3.1.1.59 Juvenile-hormone esterase 1e-34
[EC]        3.1.1.3 Triacylglycerol lipase 3e-52
[PIRKW]     duplication 2e-47
[PIRKW]     homotetramer 3e-67
[PIRKW]     transmembrane protein 9e-44
[PIRKW]     microsome 1e-130
[PIRKW]     pancreas 3e-52
[PIRKW]     endoplasmic reticulum 1e-134
[PIRKW]     homotrimer 1e-134
[PIRKW]     phosphatidylinositol linkage 5e-7?
[PIRKW]     synapse 3e-73
[PIRKW]     liver 1e-131
[PIRKW]     heparin binding 3e-52
```

```
[PIRKW]         phosphoprotein 7e-25
[PIRKW]         glycoprotein 1e-134
[PIRKW]         thyroid hormone biosynthesis 2e-47
[PIRKW]         carboxylic ester hydrolase 0.0
[PIRKW]         monomer 2e-42
[PIRKW]         disulfide bond 2e-31
[PIRKW]         mammary gland 3e-52
[PIRKW]         alternative splicing 5e-74
[PIRKW]         iodine 2e-47
[PIRKW]         pyroglutamic acid 6e-39
[PIRKW]         hydrolase 1e-135
[PIRKW]         muscle 3e-73
[PIRKW]         thyroid gland 2e-47
[PIRKW]         membrane protein 3e-73
[PIRKW]         neurotransmitter degradation 3e-73
[PIRKW]         cholesterol 3e-52
[PIRKW]         homodimer 2e-47
[PIRKW]         nerve 3e-73
[SUPFAM]        cholinesterase 0.0
[SUPFAM]        triacylglycerol lipase 1e-32
[SUPFAM]        cholinesterase homology 0.0
[SUPFAM]        thyroglobulin 2e-47
[SUPFAM]        thyroglobulin type I repeat homology 2e-47
[SUPFAM]        juvenile-hormone esterase 2e-35
[SUPFAM]        probable lipolytic protein ybaC 1e-07
[PROSITE]       CARBOXYLESTERASE_B_2   1
[PROSITE]       CARBOXYLESTERASE_B_1   1
[PFAM]          Carboxylesterases
[KW]            Alpha_Beta
[KW]            3D
[KW]            LOW_COMPLEXITY     3.95 %

SEQ    MTAQSRSPTTPTFPGPSQRTPLTPCPVQTPRLGKALIHCWTDPGQPLGEQQRVRRQRTET
SEG    .................................................xxxxxxxx...
lacj-  ............................................................

SEQ    SEPTMRLHRLRARLSAVACGLLLLLVRGQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQ
SEG    ....................xxxxx...................................
lacj-  .......................................ETTEEEECEEEEETTEE--EE

SEQ    TFLGIPFAKPPLGPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTFPS
SEG    ............................................................
lacj-  EEEEEECEETTTGGGTTTCCEECCCCCCEEECCCCCCBCCCCCCTTTTTT-HHHHHCCCC

SEQ    DSMSEDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMASLYDGSMLAALENVVVVIIQ
SEG    ............................................................
lacj-  CCBTTTTCEEEEEET--TTTTTTEEEEEEECTTTTTTCTTTTGCHHHHHHHHCCEEEECC

SEQ    YRLGVLGFFSTGDKHATGNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSS
SEG    ............................................................
lacj-  CCCCGGGCCCTTTTTTTCCHHHHHHHHHHHHHHHCGGGGCEEEEEEEEEEECHHHHHHHH

SEQ    LVVSPISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSEALVGCLRGKS
SEG    ............................................................
lacj-  HHHCGGGTTTTCEEEEEETTTTTTTTTTBCHHHHHHHHHHHHC-CCCCCHHHHHHHHHHCC

SEQ    KEEILAINKPFKMIPGVVDGVFLPRHPQELLASADFQPVPSIVGVNNNEFGWLIPKVMRI
SEG    ............................................................
lacj-  HHHHHHHHTCCCTTTCBTTTTTTTTTHHHHHHHTTTCCCCEEEEEETBTHHHHHHTTTTT

SEQ    YDTQKEMDREASQAALQKMLTLLMLPPTFGDLLREEYIGDNGDPQTLQAQFQEMMADSMF
SEG    ............................................................
lacj-  TTTCCCCCHHHHHHHHHHHTTTTCHHHHHHHHHHCTTTTTTTHHHH-HHHHHHHHHHHHH

SEQ    VIPALQVAHFQCSRAPVYFYEFQHQPSWLKNIRPPHMKADHVKFTEEEEQLSRKMMKYWA
SEG    ............................................................
lacj-  HHHHHHHHHHHHCCCCEEEEEEECCCCGGGTTBTTTHHHCGGGCCCHHHHHHHHHHHHHH

SEQ    NFARNGNPNGEGLPHWPLFDQEEQYLQLNLQPAVGRALKAHRLQFWKKALPQKIQELEEP
SEG    .......................................................xxxxx
lacj-  HHHHHCCCCCCC--CCCCBTTTTBEEEECCCCCEEETTTHHHHHHHHHHHHH........

SEQ    EERHTEL
SEG    xxxxxx.
lacj-  .......
```

Prosite for DKFZphtes3_35n9.3

```
PS00122     279->295   CARBOXYLESTERASE_B_1   PDOC00112
PS00941     185->196   CARBOXYLESTERASE_B_2   PDOC00112
```

Pfam for DKFZphtes3_35n9.3

```
HMM_NAME     Carboxylesterases

HMM              *MfMnwlimFLLwmItWIi.WheqaprpPdPyiVdtnnCGkIRGmNedtD
                   +    +L+++ ++++++++  ++Q++++P   I  T+  G + G ++ +
Query        69   RLRARLSAVACGLLLLLVRGQGQDSASP---IRTTHT-GQVLGSLVHVK   113

HMM              NG..pYYvFlGIPYAEPPVGNLRFKePQPYhePWtNVWNATnYPPMCMQW
                 +     + +FLGIP+A+PP+G LRF +P+P +E W++V++ T+ P MC+Q+
Query       114  GANAGVQTFLGIPFAKPPLGPLRFAPPEP-PESWSGVRDGTTHPAMCLQD   162

HMM              ndFGFWlFdmieMWNeniP..eMSEDCLYLNVWTPWnrkPNskLPVMVWI
                       +++   +++N++ P   +MSEDCLYL+++TP+ + ++S+LPVMVWI
Query       163  LTAV--ESEFLSQFNMTFPSDSMSEDCLYLSIYTPAHSHEGSNLPVMVWI   210

HMM              HGGGFMFGSGhsYPliqYDgeylMMeeNVIVVtINYRLGPFGFLSTgDid
                 HGG+++FG  +      ++YDG+ L++ ENV+VV I+YRLG++GF+STGD +
Query       211  HGGALVFGMA-----SLYDGSMLAALENVVVVIIQYRLGVLGFFSTGDKH   255

HMM              lPPHGNWGLWDQRMALQWVQDNIAnFGGDPNNITIFGESAGGMSVHlHML
                 +    GNWG++DQ++AL+WVQ+NIA+FGG+P+++TIFGESAGG+SV+ ++
Query       256  AT--GNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTSVSSLVV   303

HMM              SYGGDNPPmfKqLFHRAIMQSGsAmcPWvIQsnyNaRqRAfRFArimGCN
                 S        P + +LFH AIM+SG A+ P++I S++     +    +A++  C+
Query       304  S------PISQGLFHGAIMESGVALLPGLIASSA--DVISTVVANLSACD   345

HMM              rmDssEMIqCLRsKPwEELWdAtWnFWmWfYfPFlPWFFgPVIDGDDaPE
                 + DS++++ CLR K+ EE++++++     +F  + +         +DG+
Query       346  QVDSEALVGCLRGKSKEEILAINK----PFKMIPGV------VDGV----   381

HMM              aFIPDHPeeMIkEGkFnDVPWIIGYNnDEGiWFapMmMnfnWfdEDeWId
                  F+P+HP+E++++  F  VP I+G+NN E++W++P  M  + +  +E++
Query       382  -FLPRHPQELLASADFQPVPSIVGVNNNEFGWLIPKVMRIYDT-QKEMDR   429

HMM              itNedWyeWMPYIlFYrddmsNikDMDDYiDkvyEeYPgWWDrFPqESYW
                 ++ +  ++ M  +L  + + +          D ++EEY+G+  + PQ
Query       430  EASQAALQKMLTLLMLPPT-F--------GDLLREEYIGDNGD-PQTLQA   469

HMM              nLqDMFTDYLFWCPtRihadnHRkHwgsPVYMYeFDHPpSFGYgQFFmWR
                 ++Q+M+ D  F++P   + ++H++  +PVY+YEF+H PS           +
Query       470  QFQEMMADSMFVIP--ALQVAHFQCSRAPVYFYEFQHQPSW------LKN   511

HMM              WWPpWMgvdH*
                  +PP+M++DH
Query       512  IRPPHMKADH      521

HMM              *tEEEiissMRmMMNYWINFAKhGNPNnthnglCWWPqYTsnEQYdMIMe
                  TEEE+ +S R MM+YW+NFA++GNPN++  GL++WP ++++EQY++  +
Query       525   TEEEQLS-RKMMKYWANFARNGNPNGE--GLPHWPLFDQEEQYLQLNL   570

HMM              tIImiQmCrmrDPYCNFW*
                 +         +++++  + FW
Query       571  QPAVGRALKAHR--LQFW     586
```

DKFZphtes3_35p17
----

group: testes derived

DKFZphtes3_35p17 encodes a novel 505 amino acid protein with weak similarity to Proteins of the armadillo family.

Proteins of the armadillo family are involved in diverse cellular processes in higher eukaryotes. Some of them, like armadillo, beta-catenin and plakoglobins have dual functions in intercellular junctions and signalling cascades. Others, belonging to the importin-alpha-subfamily are involved in NLS recognition and nuclear transport, while some members of the armadillo family have as yet unknown functions. The novel protein shows similarity to S. cerevisiae protein Yel013p (VAC8) and Danio rerio b-catenin, but contains no armadillo (arm) repeats.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to S.cerevisiae VAC8

complete cDNA, complete cds, few EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1966 bp
Poly A stretch at pos. 1956, polyadenylation signal at pos. 1935

```
   1 AAGTCAAATG TAAGATTGGT TCATTAAAAA TACTGAAGGA AATCAGTCAT
  51 AATCCTCAAA TCAGACAGAA TATTGTTGAC CTTGGGGGCT TACCAATTAT
 101 GGTGAATATA CTTGATTCTC CACACAAGAG TCTAAAATGT TTGGCAGCCG
 151 AGACTATCGC GAATGTTGCC AAGTTTAAAA GAGCACGGCG GGTGGTGAGG
 201 CAGCACGGGG GTATCACCAA ACTGGTTGCT CTACTAGACT GTGCACATGA
 251 TTCCACAAAA CCTGCCCAAT CGAGTCTGTA TGAGGCCAGA GACGTGGAAG
 301 TGGCTCGCTG TGGGGCACTG GCCCTGTGGA GCTGCAGTAA GAGTCATACG
 351 AATAAAGAAG CCATCCGCAA AGCTGGGGGC ATTCCTCTGT TGGCTCGGCT
 401 GCTGAAGACT TCTCATGAAA ACATGCTAAT TCCAGTGGTG GGGACATTGC
 451 AAGAGTGTGC ATCAGAGGAA AACTACCGGG CTGCAATCAA AGCAGAAAGG
 501 ATCATTGAAA ACCTTGTCAA GAACCTAAAT AGTGAGAATG AGCAGCTGCA
 551 GGAGCACTGC GCCATGGCCA TTTACCAGTG TGCTGAAGAT AAGGAAACCC
 601 GGGACCTCGT TAGGCTGCAC GGAGGACTTA AGCCCTTGGC CAGTCTACTC
 651 AATAACACTG ACAATAAAGA GCGGTTAGCT GCTGTCACAG GGGCTATATG
 701 GAAATGTTCC ATCAGCAAAG AGAATGTTAC CAAGTTTCGG GAATACAAAG
 751 CCATTGAAAC CTTGGTGGGA CTTCTAACAG ATCAGCCTGA AGAAGTACTT
 801 GTGAATGTGG TTGGGGCCTT GGGAGAATGC TGCCAAGAAC GTGAAAACCG
 851 AGTCATTGTC CGGAAATGTG GTGGCATTCA ACCACTTGTG AACCTCCTTG
 901 TTGGAATAAA CCAAGCTCTT CTTGTGAATG TTACAAAAGC AGTTGGTGCT
 951 TGTGCAGTAG AACCTGAAAG TATGATGATA ATTGATCGCT TAGATGGAGT
1001 TCGTTTGTTG TGGTCCCTGC TGAAAAATCC TCACCCAGAC GTGAAGGCCA
1051 GCGCAGCATG GGCACTCTGT CCATGCATCA AAAATGCAAA GGATGCTGGG
1101 GAAATGGTTC GTTCCTTTGT TGGTGGTTTG GAACTTATTG TCAATTTACT
1151 GAAATCAGAT AACAAAGAAG TTCTGGCAAG TGTATGTGCT GCCATTACCA
1201 ACATAGCAAA AGATCAAGAA AATTTAGCTG TTATCACAGA TCATGGAGTT
1251 GTTCCTTTAT TGTCCAAACT GGCAAATACA AATAACAATA AATTGAGACA
1301 TCATCTAGCA GAAGCTATTT CACGTTGCTG TATGTGGGGC AGGAATAGAG
1351 TGGCCTTCGG TGAGCACAAA GCAGTGGCTC CACTAGTGCG TTATCTGAAA
1401 TCAAATGACA CCAACGTGCA TCGGGCGACA GCTCAGGCCT TGTACCAACT
1451 CTCAGAAGAC GCCGATAACT GCATCACCAT GCATGAGAAT GGTGCAGTAA
1501 AGCTTCTACT GGATATGGTT GGGTCCCCTG ACCAGGATCT CCAGGAAGCT
1551 GCAGCTGGTT GTATATCCAA TATCCGCAGG CTGGCTCTTG CTACAGAGAA
1601 GGCAAGATAC ACTTGAAATT TAAATGGACA TTACAAGCTA TCAAATTCTA
1651 CATGACACAG GACATGTCAC TCCCATGGCC AGAAAGCCTA AATTGGGAAA
1701 CAGTTGTTAG CAAACCCTTT CAACCATCTA AATGAAAACA CACAAATTGA
1751 AAATGCACAG AATGTTTTTC ATCTGAAAAT TGCATGGAGA CTTTTGTTTC
1801 TATTTAATGT TTTCGAGATA TGACATGTGA TAAGATGGAA AGCCAATAAA
1851 CCTGTGATAA GTTTCTAAGA ATATGAGAAT ATACGTATAT GATGTATTTT
1901 TAGTTCAGTG ATGCTTTTGT ATTTGTGGCG ATTTTAATAA AGGATATGGC
1951 CTTCCCAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

98413148:
Yel013p (Vac8p), an armadillo repeat protein related to plakoglobin and importin alpha is associated with the yeast vacuole membrane.

98330438:
YEB3/VAC8 encodes a myristylated armadillo protein of the Saccharomyces cerevisiae vacuolar membrane that functions in vacuole fusion and inheritance.

98158703:
Vac8p, a vacuolar protein with armadillo repeats, functions in both vacuole inheritance and protein targeting from the cytoplasm to vacuole.

Peptide information for frame 3

ORF from 99 bp to 1613 bp; peptide length: 505
Category: similarity to known protein
Classification: unset

```
  1 MVNILDSPHK SLKCLAAETI ANVAKFKRAR RVVRQHGGIT KLVALLDCAH
 51 DSTKPAQSSL YEARDVEVAR CGALALWSCS KSHTNKEAIR KAGGIPLLAR
101 LLKTSHENML IPVVGTLQEC ASEENYRAAI KAERIIENLV KNLNSENEQL
151 QEHCAMAIYQ CAEDKETRDL VRLHGGLKPL ASLLNNTDNK ERLAAVTGAI
201 WKCSISKENV TKFREYKAIE TLVGLLTDQP EEVLVNVVGA LGECCQEREN
251 RVIVRKCGGI QPLVNLLVGI NQALLVNVTK AVGACAVEPE SMMIIDRLDG
301 VRLLWSLLKN PHPDVKASAA WALCPCIKNA KDAGEMVRSF VGGLELIVNL
351 LKSDNKEVLA SVCAAITNIA KDQENLAVIT DHGVVPLLSK LANTNNNKLR
401 HHLAEAISRC CMWGRNRVAF GEHKAVAPLV RYLKSNDTNV HRATAQALYQ
451 LSEDADNCIT MHENGAVKLL LDMVGSPDQD LQEAAAGCIS NIRRLALATE
501 KARYT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35p17, frame 3

PIR:S50446 VAC8 protein - yeast (Saccharomyces cerevisiae), N = 1, Score = 237, P = 7.8e-17

PIR:T00403 T13E15.9 protein - Arabidopsis thaliana, N = 1, Score = 215, P = 4.9e-14

TREMBL:DR41081_1 product: "b-catenin"; Danio rerio b-catenin mRNA, complete cds., N = 1, Score = 195, P = 5.8e-12

>PIR:S50446 VAC8 protein - yeast (Saccharomyces cerevisiae)
Length = 578

HSPs:

Score = 237 (35.6 bits), Expect = 7.8e-17, P = 7.8e-17
Identities = 106/401 (26%), Positives = 177/401 (44%)

```
Query:    92 AGGIPLLARLLKTSHENMLIPVVGTLQECASEENYRAAIKAERIIENLVKNLNSENEQLQ 151
             +GG PL A      +N+ +     L      E Y   +  E ++E ++  L S++ Q+Q
Sbjct:    45 SGG-PLKALTTLVYSDNLNLQRSAALAFAEITEKYVRQVSRE-VLEPILILLQSQDPQIQ 102

Query:   152 EHCAMAIYQCAEDKETRDLVRLHGGLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVT 211
                  A+   A + E + L+    GGL+PL + +    DN E       G I    +    +N
Sbjct:   103 VAACAALGNLAVNNENKLLIVEMGGLEPLINQMMG-DNVEVQCNAVGCITNLATRDDNKH 161

Query:   212 KFREYKAIETLVGLLTDQPEEVLVNVVGALGECCQERENRVIVRKCGGIQPLVNLLVGIN 271
             K     A+  L  L    +   V  N  GAL          ENR  +   G +  LV+LL    +
Sbjct:   162 KIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNAGAVPVLVSLLSSTD 221

Query:   272 QALLVNVTKAVGACAVEPESMMIIDRLDG--VRLLWSLLKNPHPDVKASAAWALCPCIKN 329
               +     T A+   AV+  +    + + +   V  L SL+ +P   VK  A  AL      +
```

```
Sbjct:   222 PDVQYYCTTALSNIAVDEANRKKLAQTEPRLVSKLVSLMDSPSSRVKCQATLALRNLASD 281

Query:   330 AKDAGEMVRSFVGGLELIVNLLKSDNKE-VLASVCAAITNIAKDQENLAVITDHGVV-PL 387
                 E+VR+  GGL  +V L++SD+   VLASV A I NI+     N  +I D G + PL
Sbjct:   282 TSYQLEIVRA--GGLPHLVKLIQSDSIPLVLASV-ACIRNISIHPLNEGLIVDAGFLKPL 338

Query:   388 LSKLANTNNNKLRHHLAEAISRCCMWG-RNRVAFGEHKAVAPLVRYLKSNDTNVHRATAQ 446
             +  L    ++ +++ H      +             +NR  F E  AV             +  +V ++
Sbjct:   339 VRLLDYKDSEEIQCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSV-QSEIS 397

Query:   447 ALYQLSEDAD-NCITMHENGAVKLLLDMVGSPDQDLQEAAAGCISNI 492
             A + +   AD + + + E   +  L+ M  S +Q++    AA  ++N+
Sbjct:   398 ACFAILALADVSKLDLLEANILDALIPMTFSQNQEVSGNAAAALANL 444

 Score = 213 (32.0 bits), Expect = 3.6e-14, P = 3.6e-14
 Identities = 81/341 (23%), Positives = 163/341 (47%)

Query:   163 EDKETRDLVRLHGGLKPLASLLNNTD-NKERLAAVTGAIWKCSISKENVTKFREYKAIET 221
             EDK+  D     G LK L +L+ + + + N +R AA+  A      I+++ V +     + +E
Sbjct:    36 EDKDQLDFYS-GGPLKALTTLVYSDNLNLQRSAALAFA----EITEKYVRQVSR-EVLEP 89

Query:   222 LVGLLTDQPEEVLVNVVGALGECCQERENRVIVRKCGGIQPLVNLLVGINQALLVNVTKA 281
             ++ LL  Q  ++ V    ALG     EN++++ + GG++PL+N ++G N  +  N
Sbjct:    90 ILILLQSQDPQIQVAACAALGNLAVNNENKLLIVEMGGLEPLINQMMGDNVEVQCNAVGC 149

Query:   282 VGACAVEPESMMIIDRLDGVRLLWSLLKNPHPDVKASAAWALCPCIKNAKDAGEMVRSFV 341
             +  A  ++  I    + L  L K+ H  V+ +A  AL     + ++  E+V +
Sbjct:   150 ITNLATRDDNKHKIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNA-- 207

Query:   342 GGLELIVNLLKSDNKEVLASVCAAITNIAKDQENLAVI--TDHGVVPLLSKLANTNNNKL 399
             G + ++V+LL  S + +V    A++NIA D+ N  + T+ +V  L  L ++ ++++
Sbjct:   208 GAVPVLVSLLSSTDPDVQYYCTTALSNIAVDEANRKKLAQTEPRLVSKLVSLMDSPSSRV 267

Query:   400 RHHLAEAISRCCMWGRNRVAFGEHKAVAPLVRYLKSNDTNVHRATAQALYQLSEDADNCI 459
             +    A+         ++    +  LV+ ++S+  +  A+  +  +S   N
Sbjct:   268 KCQATLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHPLNEG 327

Query:   460 TMHENGAVKLLLDMVGSPDQDLQEAAAGCISNIRRLALATEKAR 503
             + + G +K L+ ++   D + E   +S +R LA ++EK R
Sbjct:   328 LIVDAGFLKPLVRLLDYKDSE--EIQCHAVSTLRNLAASSEKNR 369

 Score = 180 (27.0 bits), Expect = 1.6e-10, P = 1.6e-10
 Identities = 80/346 (23%), Positives = 142/346 (41%)

Query:   145 SENEQLQEHCAMAIYQCAEDKETRDLVRLHGGLKPLASLLNNTDNKERLAAVTGAIWKCS 204
             S+N  LQ   A+A  +  E K  R + R    L+P+ LL + D +  ++AA  A+ +
Sbjct:    58 SDNLNLQRSAALAFAEITE-KYVRQVSR--EVLEPILILLQSQDPQIQVAACA-ALGNLA 113

Query:   205 ISKENVTKFREYKAIETLVGLLTDQPEEVLVNVVGALGECCQERENRVIVRKCGGIQPLV 264
             ++ EN       E +E L+  +    EV  N VG +         +N+ +  G + PL
Sbjct:   114 VNNENKLLIVEMGGLEPLINQMMGDNVEVQCNAVGCITNLATRDDNKHKIATSGALIPLT 173

Query:   265 NLLVGINQALLVNVTKAVGACAVEPESMMIIDRLDGVRLLWSLLKNPHPDVKASAAWALC 324
             L    +  +  N T A+    E+    +   V +L SLL +  PDV+    AL
Sbjct:   174 KLAKSKHIRVQRNATGALLNMTHSEENRKELVNAGAVPVLVSLLSSTDPDVQYYCTTALS 233

Query:   325 PCIKNAKDAGEMVRSFVGGLELIVNLLKSDNKEVLASVCAAITNIAKDQENLAVITDHGV 384
             +  +  ++  ++    +V+L+ S  +  V  A+ N+A D  I  G
Sbjct:   234 NIAVDEANRKKLAQTEPRLVSKLVSLMDSPSSRVKCQATLALRNLASDTSYQLEIVRAGG 293

Query:   385 VPLLSKLANTNNNKLRHHLAEAISRCCMWGRNRVAFGEHKAVAPLVRYLKSNDTNVHRAT 444
             +P L KL  +++ L   I   +  N  +  + PLVR L  D+  +
Sbjct:   294 LPHLVKLIQSDSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPLVRLLDYKDSEEIQCH 353

Query:   445 A-QALYQLSEDAD-NCITMHENGAVKLLLDMVGSPDQDLQEAAAGCIS 490
             A  L L+ ++ N  E+GAV+ ++  +Q  + C +
Sbjct:   354 AVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSVQSEISACFA 401

 Score = 155 (23.3 bits), Expect = 8.8e-08, P = 8.8e-08
 Identities = 88/401 (21%), Positives = 175/401 (43%)

Query:    60 LYEARD--VEVARCGALALWSCSKSHTNKEAIRKAGGI-PLLARLLKTSHENMLIPVVGT 116
             L +++D  ++VA C AL  + + ++ NK  I + GG+ PL+ +++ + E +  VG
Sbjct:    93 LLQSQDPQIQVAACAALG--NLAVNNENKLLIVEMGGLEPLINQMMGDNVE-VQCNAVGC 149

Query:   117 LQECASEENYRAAIKAERIIENLVKNLNSENEQLQEHCAMAIYQCAEDKETR-DLVRLHG 175
             +  A+ ++ + I  + L K S++ ++Q + A+ +E R +LV G
Sbjct:   150 ITNLATRDDNKHKIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNA-G 208

Query:   176 GLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVTKFR--EYKAIETLVGLLTDQPEEV 233
             + L SLL++TD + T A+ ++ + N K E + + LV L+ V
```

```
Sbjct:   209 AVPVLVSLLSSTDPDVQYYCTT-ALSNIAVDEANRKKLAQTEPRLVSKLVSLMDSPSSRV 267

Query:   234 LVNVVGALGECCQERENRVIVRKCGGIQPLVNLLVGINQALLVNVTKAVGACAVEPESMM 293
                  AL      +    ++ + + GG+  LV L+    +  L++       +    ++ P +
Sbjct:   268 KCQATLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHPLNEG 327

Query:   294 IIDRLDGVRLLWSLLK-NPHPDVKASAAWALCPCIKNA-KDAGEMVRSFVGGLELIVNLL 351
             +I     ++ L  LL       +++  A    L       ++ K+  E   S  G +E    L
Sbjct:   328 LIVDAGFLKPLVRLLDYKDSEEIQCHAVSTLRNLAASSEKNRKEFFES--GAVEKCKELA 385

Query:   352 KSDNKEVLA--SVCAAITNIAKDQENLAVITDHGVVPLLSKLANTNNNKLRHHLAEAISR 409
                  V +   S C AI  +A D   L ++ +   ++  L  +  + N ++  + A A++
Sbjct:   386 LDSPVSVQSEISACFAILALA-DVSKLDLL-EANILDALIPMTFSQNQEVSGNAAAALAN 443

Query:   410 CCMWGRNRVAFGE-----HKAVAP-LVRYLKSNDTNVHRATAQALYQLSE 453
              C     N      E        ++ +   L+R+LKS+              + QL E
Sbjct:   444 LCSRVNNYTKIIEAWDRPNEGIRGFLIRFLKSDYATFEHIALWTILQLLE 493

 Score = 139 (20.9 bits), Expect = 5.0e-06, P = 5.0e-06
 Identities = 80/329 (24%), Positives = 142/329 (43%)

Query:    37 GGITKLVALLDCAHD-STKPAQ---SSLYEARDVEVARCGALALWSCSKSHTNKEAIRKA 92
             G IT L     D  H  +T  A     + L +++ + V R    AL  + +  S   N++  +  A
Sbjct:   148 GCITNLATRDDNKHKIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNA 207

Query:    93 GGIPLLARLLKTSHENMLIPVVGTLQECASEE-NYRAAIKAE-RIIENLVKNLNSENEQL 150
             G +P+L  LL ++   ++          L  A +E N +   + E R++ LV  ++S + ++
Sbjct:   208 GAVPVLVSLLSSTDPDVQYYCTTALSNIAVDEANRKKLAQTEPRLVSKLVSLMDSPSSRV 267

Query:   151 QEHCAMAIYQCAEDKETR-DLVRLHGGLKPLASLLNNTDNKERLAAVTGAIWKCSISKEN 209
             +     +A++    A D   + ++VR  GGL  L  L+ + D+   +  A   I    SI   N
Sbjct:   268 KCQATLALRNLASDTSYQLEIVRA-GGLPHLVKLIQS-DSIPLVLASVACIRNISIHPLN 325

Query:   210 VTKFREYKAIETLVGLLT-DQPEEVLVNVVGALGECCQERE-NRVIVRKCGGIQPLVNLL 267
                 +    ++ LV LL     EE+  + V  L        E NR    + G  ++    L
Sbjct:   326 EGLIVDAGFLKPLVRLLDYKDSEEIQCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELA 385

Query:   268 VG--INQALLVNVTKAVGACA-VEPESMMIIDRLDGVRLLWSLLKNPHPDVKASAAWA-L 323
             +    ++     ++      A+ A A V     ++ + LD + + +  +N   A+AA A L
Sbjct:   386 LDSPVSVQSEISACFAILALADVSKLDLLEANILDAL-IPMTFSQNQEVSGNAAAALANL 444

Query:   324 CPCIKN-AKDAGEMVRSFVGGLELIVNLLKSD 354
             C  + N  K        R  G      ++  LKSD
Sbjct:   445 CSRVNNYTKIIEAWDRPNEGIRGFLIRFLKSD 476

 Score = 136 (20.4 bits), Expect = 1.1e-05, P = 1.1e-05
 Identities = 72/304 (23%), Positives = 133/304 (43%)

Query:    58 SSLYEARDVEVARCGALALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHENMLIPVVGTL 117
             + L +++ + V R    AL  + +  S   N++  +  AG +P+L  LL ++   ++       L
Sbjct:   173 TKLAKSKHIRVQRNATGALLNMTHSEENRKELVNAGAVPVLVSLLSSTDPDVQYYCTTAL 232

Query:   118 QECASEE-NYRAAIKAE-RIIENLVKNLNSENEQLQEHCAMAIYQCAEDKETR-DLVRLH 174
                  A +E N +   + E R++ LV  ++S + ++ +++     +A++    A D   + ++VR
Sbjct:   233 SNIAVDEANRKKLAQTEPRLVSKLVSLMDSPSSRVKCQATLALRNLASDTSYQLEIVRA- 291

Query:   175 GGLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLT-DQPEEV 233
             GGL  L  L+ + D+   +  A   I    SI   N    +    ++ LV LL     EE+
Sbjct:   292 GGLPHLVKLIQS-DSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPLVRLLDYKDSEEI 350

Query:   234 LVNVVGALGECCQERE-NRVIVRKCGGIQPLVNLLVG--INQALLVNVTKAVGACA-VEP 289
              + V  L        E NR    + G ++    L   ++     ++      A+ A A V
Sbjct:   351 QCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSVQSEISACFAILALADVSK 410

Query:   290 ESMMIIDRLDGVRLLWSLLKNPHPDVKASAAWA-LCPCIKN-AKDAGEMVRSFVGGLELI 347
              ++   + LD + + +  +N     A+AA A LC  + N  K        R  G      +
Sbjct:   411 LDLLEANILDAL-IPMTFSQNQEVSGNAAAALANLCSRVNNYTKIIEAWDRPNEGIRGFL 469

Query:   348 VNLLKSD 354
              +  LKSD
Sbjct:   470 IRFLKSD 476

 Score = 114 (17.1 bits), Expect = 2.7e-03, P = 2.7e-03
 Identities = 71/335 (21%), Positives = 132/335 (39%)

Query:     1 MVNILDSPHKSLKCLAAETIANVAKFKRARRVVRQHGGITKLVALLDCAHDSTKPAQSSL 60
             +  +  S H  ++  A    + N+    + R+ +   G +  LV++LL         ST P
Sbjct:   172 LTKLAKSKHIRVQRNATGALLNMTHSEENRKELVNAGAVPVLVSLLS----STDP----- 222

Query:    61 YEARDVEVARCGALALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHENMLIPVVGTLQEC 120
                 DV+      AL+  +   +++  K A   +   L   L+ +   +         L+
```

```
Sbjct:   223 ----DVQYYCTTALSNIAVDEANRKKLAQTEPRLVSKLVSLMDSPSSRVKCQATLALRNL 278

Query:   121 ASEENYRAAIKAERIIENLVKNLNSENEQLQEHCAMAIYQCAEDKETRDLVRLHGGLKPL 180
             AS+ +Y+  I       + +LVK + S++  L         I    +       L+   G LKPL
Sbjct:   279 ASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPL 338

Query:   181 ASLLNNTDNKERLAAVTGAIWKCSISKE-NVTKFREYKAIETLVGLLTDQPEEVLVNVVG 239
               LL+  D++E            +  + S E N  +F E  A+E    L  D P  V   +
Sbjct:   339 VRLLDYKDSEEIQCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSVQSEISA 398

Query:   240 ALGECCQERENRVIVRKCGGIQPLVNLLVGINQALLVNVTKAVG-ACAVEPESMMIIDRL 298
                   +++ + +    +   L+ +      NQ +  N     A+    C+        II+
Sbjct:   399 CFAILALADVSKLDLLEANILDALIPMTFSQNQEVSGNAAAALANLCSRVNNYTKIIEAW 458

Query:   299 D----GVR-LLWSLLKNPHPDVKASAAWALCPCIKNAKDAGE 335
             D    G+R  L   LK+ +   +  A W +   +++  D  E
Sbjct:   459 DRPNEGIRGFLIRFLKSDYATFEHIALWTILQLLESHNDKVE 500

 Score = 106 (15.9 bits), Expect = 2.0e-02, P = 2.0e-02
 Identities = 49/204 (24%), Positives = 89/204 (43%)

Query:    65 DVEVARCGALA-LWSCSKSHTNKEAIRKAGGIPLLARLLKTSHENMLIPVVGTLQECA-S 122
             +VEV +C A+  + + +   NK  I  +G +  L +L K+ H  +       G L     S
Sbjct:   139 NVEV-QCNAVGCITNLATRDDNKHKIATSGALIPLTKLAKSKHIRVQRNATGALLNMTHS 197

Query:   123 EENYRAAIKAERIIENLVKNLNSENEQLQEHCAMAIYQCAEDKETRD-LVRLHGGL-KPL 180
             EEN +   +  A  +    LV  L+S +  +Q +C  A+    A D+  R  L +      L    L
Sbjct:   198 EENRKELVNAGAV-PVLVSLLSSTDPDVQYYCTTALSNIAVDEANRKKLAQTEPRLVSKL 256

Query:   181 ASLLNNTDNKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLTDQPEEVLVNVVGA 240
             SL+++  ++ + A T A+   +           +   +       LV L+     +++  V
Sbjct:   257 VSLMDSPSSRVKCQA-TLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVAC 315

Query:   241 LGECCQERENRVIVRKCGGIQPLVNLL 267
             +         N  ++   G ++PLV LL
Sbjct:   316 IRNISIHPLNEGLIVDAGFLKPLVRLL 342
```

Pedant information for DKFZphtes3_35p17, frame 3

Report for DKFZphtes3_35p17.3

```
[LENGTH]    505
[MW]        55224.34
[pI]        8.43
[HOMOL]     PIR:S50446 VAC8 protein - yeast (Saccharomyces cerevisiae) 2e-16
[FUNCAT]    30.25 vacuolar and lysosomal organization   [S. cerevisiae, YEL013w] 8e-18
[FUNCAT]    06.04 protein targeting, sorting and translocation   [S. cerevisiae, YEL013w]
8e-18
[FUNCAT]    09.25 vacuolar and lysosomal biogenesis   [S. cerevisiae, YEL013w] 8e-18
[FUNCAT]    08.01 nuclear transport   [S. cerevisiae, YNL189w] 3e-06
[FUNCAT]    03.22 cell cycle control and mitosis [S. cerevisiae, YNL189w] 3e-06
[FUNCAT]    30.10 nuclear organization   [S. cerevisiae, YNL189w] 3e-06
[BLOCKS]    BL01265C
[BLOCKS]    BL00242A Integrins alpha chain proteins
[SCOP]      d3bct__ 1.91.1.1.1 beta-Catenin [Mouse (Mus musculus) 7e-18
[PIRKW]     cytosol 3e-11
[PIRKW]     apoptosis 3e-11
[PIRKW]     carcinogenesis 3e-11
[PIRKW]     cell adhesion 3e-11
[PIRKW]     cytoskeleton 3e-12
[SUPFAM]    pendulin 1e-07
[KW]        All_Alpha
[KW]        3D
[KW]        LOW_COMPLEXITY    2.38 %

SEQ    MVNILDSPHKSLKCLAAETIANVAKFKRARRVVRQHGGITKLVALLDCAHDSTKPAQSSL
SEG    .....................xxxxxxxxxxx.........................
2bct-  ..........................................................HH

SEQ    YEARDVEVARCGALALWSCSKSHTNKEAIRKAGGIPLLARLLKTSHENMLIPVVGTLQEC
SEG    ............................................................
2bct-  HHCCCHHHHHHHHHHHHHHHHHCHHHHHHHHHCCHHHHHHHGGGCCCHHHHHHHHHHHHH

SEQ    ASEENYRAAIKAERIIENLVKNLNSENEQLQEHCAMAIYQCAEDKETRDLVRLHGGLKPL
SEG    ............................................................
2bct-  HHTTTHHHHHHHHCHHHHHHHHHCCCCHHHHHHHHHHHHHHTTHHHHHHHHHHCHHHHH
```

```
SEQ     ASLLNNTDNKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLTDQPEEVLVNVVGA
SEG     ............................................................
2bct-   HHHHH-HCCCHHHHHHHHHHHHHHHHCCCHHHHHHHHHHCHHHHHHTTTTTCCHHHHHHHHHH

SEQ     LGECCQERENRVIVRKCGGIQPLVNLLVGINQALLVNVTKAVGACAVEPESMMIIDRLDG
SEG     ............................................................
2bct-   H------HHHHHCCCCTTTHHHHHHHHHHHHCTTTHHHHHHHHHHTTTHHHHHHHH-HHCH

SEQ     VRLLWSLLKNPHPDVKASAAWALCPCIKNAKDAGEMVRSFVGGLELIVNLLKSDNKEVLA
SEG     ............................................................
2bct-   HHHHHHHHHTTTHHHHHHHHHHHHHHHHCCCCHH-HHHHHHHHHHHHHHHHHCTTTTTHHHH

SEQ     SVCAAITNIAKDQENLAVITDHGVVPLLSKLANTNNNKLRHHLAEAISRCCMWGRNRVAF
SEG     ............................................................
2bct-   HHHHHHHHHHHCGGGHHHHHHHCHHHHHHHHHHHHHHHTTTCCHHHHHHHHHHHHCHHHHH

SEQ     GEHKAVAPLVRYLKSNDTNVHRATAQALYQLSEDADNCITMHENGAVKLLLDMVGSPDQD
SEG     ............................................................
2bct-   HTTTHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHTTHHHHHHHHHHCCHHHHHHHHTTTTTTHH

SEQ     LQEAAAGCISNIRRLALATEKARYT
SEG     .........................
2bct-   HHHHHHHHH................
```

(No Prosite data available for DKFZphtes3_35p17.3)

(No Pfam data available for DKFZphtes3_35p17.3)

DKFZphtes3_35p22
----

group: cell cycle

DKFZphtes3_35p22 encodes a novel 549 amino acid protein, with similarity to oncogene 1 (tre-2 locus).

The novel protein is closely raleted to human tre-2 and other enzymes involved in the degradation of ubiquitinated proteins. The human tre-2 oncogene encodes a deubiquitinating enzyme, indicating a role for the ubiquitin system in mammalian growth control.

The novel protein can find application in cancer diagnostics and treatment, and in regulating protein stability and growth control via regulation of ubiquitination.

strong similarity to oncogene 1 (tre-2 locus)

membrane regions: 1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: map="17"

Insert length: 2072 bp
Poly A stretch at pos. 2062, polyadenylation signal at pos. 2039

```
   1 GTTACACACA GGCAGTGGTA TCTGTGAGCA GCTCTGTGGA CTCAAAGGTT
  51 TTCTCCCTGA GAGGCATGAC CCAGGCCAGC TGATTCATCA GAATCAGGAT
 101 GGACGTGGTA GAGGTCGCGG GCAGTTGGTG GGCACAAGAG CGAGAGGACA
 151 TCATTATGAA ATACGAAAAG GGACACCGAG CTGGGCTGCC AGAGGACAAG
 201 GGGCCTAAGC CTTTTCGAAG CTACAACAAC AACGTCGATC ATTTGGGGAT
 251 TGTACATGAG ACGGAGCTGC CTCCTCTGAC TGCGCGGGAG GCGAAGCAAA
 301 TTCGGCGGGA GATCAGCCGA AAGAGCAAGT GGGTGGATAT GCTGGGAGAC
 351 TGGGAGAAAT ACAAAAGCAG CAGAAAGCTC ATAGATCGAG CGTACAAGGG
 401 AATGCCCATG AACATCCGGG GCCCGATGTG GTCAGTCCTC CTGAACACTG
 451 AGGAAATGAA GTTGAAAAAC CCCGGAAGAT ACCAGATCAT GAAGGAGAAG
 501 GGCAAGAAGT CATCTGAGCA CATCCAGCGC ATCGACCGGG ACGTAAGCGG
 551 GACATTAAGG AAGCATATAT TCTTCAGGGA TCGATACGGA ACCAAGCAGC
 601 GGGAACTACT CCACATCCTC CTGGCATATG AGGAGTACAA CCCGGAGGTG
 651 GGCTACTGCA GGGACCTGAG CCACATCGCC GCCTTGTTCC TCCTCTATCT
 701 TCCTGAGGAG GATGCATTCT GGGCACTGGT GCAGCTGCTG GCCAGTGAGA
 751 GGCACTCCCT GCAGGGATTT CACAGCCCAA ATGGCGGGAC CGTCCAGGGG
 801 CTCCAAGACC AACAGGAGCA TGTGGTAGCC ACGTCACAAC CCAAGACCAT
 851 GGGGCATCAG GACAAGAAAG ATCTATGTGG GCAGTGTTCC CCGTTAGGCT
 901 GCCTCATCCG GATATTGATT GACGGGATCT CTCTCGGGCT CACCCTGCGC
 951 CTGTGGGACG TGTATCTGGT AGAAGGCGAA CAGGCGCTGA TGCCGATAAC
1001 AAGAATCGCC TTTAAGGTTC AGCAGAAGCG CCTCACGAAG ACGTCCAGGT
1051 GTGGCCCGTG GGCACGTTTT TGCAACCGGT TCGTTGATAC CTGGGCCAGG
1101 GATGAGGACA CTGTGCTCAA GCATCTTAGG GCCTCTATGA AGAAACTAAC
1151 AAGAAAGAAG GGGGACCTGC CACCCCCAGC CAAACCCGAG CAAGGGTCGT
1201 CGGCATCCAG GCCTGTGCCG GCTTCACGTG GCGGGAAGAC CCTCTGCAAG
1251 GGGGACAGGC AGGCCCCTCC AGGCCCACCA GCCCGGTTCC CGCGGCCCAT
1301 TTGGTCAGCT TCCCCGCCAC GGGCACCTCG TTCTTCCACA CCCTGTCCTG
1351 GTGGGGCTGT CCGGGAAGAC ACCTACCCTG TGGGCACTCA GGGTGTGCCC
1401 AGCCCGGCCC TGGCTCAGGG AGGACCTCAG GGTTCCTGGA GATTCCTGCA
1451 GTGGAACTCC ATGCCCCGCC TCCCAACGGA CCTGGACGTA GAGGGCCCTT
1501 GGTTCCGCCA TTATGATTTC AGACAGAGCT GCTGGGTCCG TGCCATATCC
1551 CAGGAGGACC AGCTGGCCCC CTGCTGGCAG GCTGAACACC CTGCGGAGCG
1601 GGTGAGATCG GCTTTCGCTG CACCCAGCAC TGATTCCGAC CAGGGCACCC
1651 CCTTCAGAGC TAGGGACGAA CAGCAGTGTG CTCCCACCTC AGGGCCTTGC
1701 CTCTGCGGCC TCCACTTGGA AAGTTCTCAG TTCCCTCCAG GCTTCTAGAA
1751 GCATCTGGGC CAGGGCTCAT GGCTGGATAA TTTCCCTAGG CTTAACAACC
1801 CAAGCAAGCT TCGCATCCTC GTTTTATTTT TGGTTAAACT TATGAAAATG
1851 TATTAAGAAA GAGTGCAGCT CGAGAGAGAT TCAGAGATGG AACACACCAG
1901 ACCCCAGATC ACAAAGCCAA CCATGCCCAG CCCCTCCCAG CACCCCCAGC
1951 CCCACGACCA TCGTTCTGAA TTCTGACGAC ACCGTGAGCC TGCCTTTGTA
2001 CTTCAAACTC ATGGAAGGAT AACCACCTTC ATGTTTTGAA ATAAATGTTT
2051 CCTGTTGAAA TGAAAAAAAA AA
```

BLAST Results

Entry AC003976 from database EMBL:
Homo sapiens chromosome 17, clone hCIT.91_J_4, complete sequence.
Score = 4385, P = 0.0e+00, identities = 881/886

14 exons

Entry HSG19723 from database EMBL:
human STS A001W35.
Score = 850, P = 1.9e-32, identities = 170/170

Medline entries

92228503:
A novel transcriptional unit of the tre oncogene widely expressed in human cancer cells.

94067315:
The yeast DOA4 gene encodes a deubiquitinating enzyme related to a product of the human tre-2 oncogene.

95176708:
UBP5 encodes a putative yeast ubiquitin-specific protease that is related to the human Tre-2 oncogene product.

Peptide information for frame 3

ORF from 99 bp to 1745 bp; peptide length: 549
Category: strong similarity to known protein

```
  1 MDVVEVAGSW WAQEREDIIM KYEKGHRAGL PEDKGPKPFR SYNNNVDHLG
 51 IVHETELPPL TAREAKQIRR EISRKSKWVD MLGDWEKYKS SRKLIDRAYK
101 GMPMNIRGPM WSVLLNTEEM KLKNPGRYQI MKEKGKKSSE HIQRIDRDVS
151 GTLRKHIFFR DRYGTKQREL LHILLAYEEY NPEVGYCRDL SHIAALFLLY
201 LPEEDAFWAL VQLLASERHS LQGFHSPNGG TVQGLQDQQE HVVATSQPKT
251 MGHQDKKDLC GQCSPLGCLI RILIDGISLG LTLRLWDVYL VEGEQALMPI
301 TRIAFKVQQK RLTKTSRCGP WARFCNRFVD TWARDEDTVL KHLRASMKKL
351 TRKKGDLPPP AKPEQGSSAS RPVPASRGGK TLCKGDRQAP PGPPARFPRP
401 IWSASPPRAP RSSTPCPGGA VREDTYPVGT QGVPSPALAQ GGPQGSWRFL
451 QWNSMPRLPT DLDVEGPWFR HYDFRQSCWV RAISQEDQLA PCWQAEHPAE
501 RVRSAFAAPS TDSDQGTPFR ARDEQQCAPT SGPCLCGLHL ESSQFPPGF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_35p22, frame 3

PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human, N = 1, Score = 2181, P = 5.5e-226

PIR:S57867 oncogene 1 - human, N = 1, Score = 1536, P = 1.2e-157

>PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human
Length = 786

HSPs:

Score = 2181 (327.2 bits), Expect = 5.5e-226, P = 5.5e-226
Identities = 405/500 (81%), Positives = 440/500 (88%)

```
Query:     1 MDVVEVAGSWWAQEREDIIMKYEKGHRAGLPEDKGPKPFRSYNNNVDHLGIVHETELPPL 60
             MD+VE A S  AQER+DI+MKY+KGHRAGLPEDKGP+P    N+++D  GI+HETELPP+
Sbjct:     1 MDMVENADSLQAQERKDILMKYDKGHRAGLPEDKGPEPV-GINSSIDRFGILHETELPPV 59

Query:    61 TAREAKQIRREISRKSKWVDMLGDWEKYKSSRKLIDRAYKGMPMNIRGPMWSVLLNTEEM 120
             TAREAK+IRRE++R SKW++MLG+WE YK S KLIDR YKG+PMNIRGP+WSVLLN +E+
Sbjct:    60 TAREAKKIRREMTRTSKWMEMLGEWETYKHSSKLIDRVYKGIPMNIRGPVWSVLLNIQEI 119

Query:   121 KLKNPGRYQIMKEKGKKSSEHIQRIDRDVSGTLRKHIFFRDRYGTKQRELLHILLAYEEY 180
             KLKNPGRYQIMKE+GK+SSEHI  ID DV  TLR H+FFRDRYG KQREL +ILLAY EY
Sbjct:   120 KLKNPGRYQIMKERGKRSSEHIHHIDLDVRTTLRNHVFFRDRYGAKQRELFYILLAYSEY 179

Query:   181 NPEVGYCRDLSHIAALFLLYLPEEDAFWALVQLLASERHSLQGFHSPNGGTVQGLQDQQE 240
             NPEVGYCRDLSHI ALFLLYLPEEDAFWALVQLLASERHSL GFHSPNGGTVQGLQDQQE
Sbjct:   180 NPEVGYCRDLSHITALFLLYLPEEDAFWALVQLLASERHSLPGFHSPNGGTVQGLQDQQE 239
```

```
Query:   241 HVVATSQPKTMGHQDKKDLCGQCSPLGCLIRILIDGISLGLTLRLWDVYLVEGEQALMPI 300
             HVV  SQPKTM HQDK+ LCGQC+ LGCL+R LIDGISLGLTLRLWDVYLVEGEQ LMPI
Sbjct:   240 HVVPKSQPKTMWHQDKEGLCGQCASLGCLLRNLIDGISLGLTLRLWDVYLVEGEQVLMPI 299

Query:   301 TRIAFKVQQKRLTKTSRCGPWARFCNRFVDTWARDEDTVLKHLRASMKKLTRKKGDLPPP 360
             T IA KVQQKRL KTSRCG WAR  N+F DTWA ++DTVLKHLRAS KKLTRK+GDLPPP
Sbjct:   300 TSIALKVQQKRLMKTSRCGLWARLRNQFFDTWAMNDDTVLKHLRASTKKLTRKQGDLPPP 359

Query:   361 AKPEQGSSASRPVPASRGGKTLCKGDRQAPPGPPARFPRPIWSASPPRAPRSSTPCPGGA 420
             AK EQGS A RPVPASRGGKTLCKG RQAPPGPPA+F RPI SASPP A R STPCPGGA
Sbjct:   360 AKREQGSLAPRPVPASRGGKTLCKGYRQAPPGPPAQFQRPICSASPPWASRFSTPCPGGA 419

Query:   421 VREDTYPVGTQGVPSPALAQGGPQGSWRFLQWNSMPRLPTDLDVEGPWFRHYDFRQSCWV 480
             VREDTYPVGTQGVPS ALAQGGPQGSWRFL+W SMPRLPTDLD+ GPWF HYDF +SCWV
Sbjct:   420 VREDTYPVGTQGVPSLALAQGGPQGSWRFLEWKSMPRLPTDLDIGGPWFPHYDFERSCWV 479

Query:   481 RAISQEDQLAPCWQAEHPAE 500
             RAISQEDQLA CWQAEH  E
Sbjct:   480 RAISQEDQLATCWQAEHCGE 499
```

Pedant information for DKFZphtes3_35p22, frame 3

Report for DKFZphtes3_35p22.3

```
[LENGTH]     549
[MW]         62159.16
[pI]         9.23
[HOMOL]      PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human 0.0
[FUNCAT]     11.01 stress response [S. cerevisiae, YGR100w] 2e-16
[FUNCAT]     04.05.01.04 transcriptional control  [S. cerevisiae, YGR100w] 2e-16
[FUNCAT]     99 unclassified proteins     [S. cerevisiae, YNL293w] 3e-15
[PIRKW]      transmembrane protein 6e-14
[PROSITE]    MYRISTYL       6
[PROSITE]    AMIDATION      1
[PROSITE]    CAMP_PHOSPHO_SITE     3
[PROSITE]    CK2_PHOSPHO_SITE      4
[PROSITE]    TYR_PHOSPHO_SITE      2
[PROSITE]    PKC_PHOSPHO_SITE      10
[KW]         TRANSMEMBRANE  1
[KW]         LOW_COMPLEXITY     5.28 %
```

```
SEQ   MDVVEVAGSWWAQEREDIIMKYEKGHRAGLPEDKGPKPFRSYNNNVDHLGIVHETELPPL
SEG   ............................................................
PRD   ccceeecchhhhhhhhhhhhhccccccccccccccceeeeecccccccccccccccccccc
MEM   ............................................................

SEQ   TAREAKQIRREISRKSKWVDMLGDWEKYKSSRKLIDRAYKGMPMNIRGPMWSVLLNTEEM
SEG   ............................................................
PRD   chhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhcccccccccceeecccccc
MEM   ............................................................

SEQ   KLKNPGRYQIMKEKGKKSSEHIQRIDRDVSGTLRKHIFFRDRYGTKQRELLHILLAYEEY
SEG   ............................................................
PRD   ccccccchhhhhhccccchhhhhhhhhhhhhcccccccccccccchhhhhhhhhhhhhhc
MEM   ............................................................

SEQ   NPEVGYCRDLSHIAALFLLYLPEEDAFWALVQLLASERHSLQGFHSPNGGTVQGLQDQQE
SEG   ............................................................
PRD   ccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccchhhhhhhhh
MEM   ............................................................

SEQ   HVVATSQPKTMGHQDKKDLCGQCSPLGCLIRILIDGISLGLTLRLWDVYLVEGEQALMPI
SEG   ............................................................
PRD   hhhhhhchhhhhhhhhcccccccchhhhhhhhccccchhhhhhhhcccccccccceeeehh
MEM   ....................MMMMMMMMMMMMMMMMMM......................

SEQ   TRIAFKVQQKRLTKTSRCGPWARFCNRFVDTWARDEDTVLKHLRASMKKLTRKKGDLPPP
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhcccchhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhcccc
MEM   ............................................................

SEQ   AKPEQGSSASRPVPASRGGKTLCKGDRQAPPGPPARFPRPIWSASPPRAPRSSTPCPGGA
SEG   ............................xxxxxxxxxxxxxxxxxxxxxxxxxxxxx...
PRD   cccccccccccccccccccccccccccceeeecccccccccccccccccccccccccccc
MEM   ............................................................
```

```
SEQ   VREDTYPVGTQGVPSPALAQGGPQGSWRFLQWNSMPRLPTDLDVEGPWFRHYDFRQSCWV
SEG   ............................................................
PRD   cccccccccccccccccccccccccceeeeeccccccccccccccccccccccccccccc
MEM   ............................................................

SEQ   RAISQEDQLAPCWQAEHPAERVRSAFAAPSTDSDQGTPFRARDEQQCAPTSGPCLCGLHL
SEG   ............................................................
PRD   cchhhhhhhhhhhhhhcchhhhhhhhcccccccccccccchhhhhccccccccccceeee
MEM   ............................................................

SEQ   ESSQFPPGF
SEG   .........
PRD   ccccccccc
MEM   .........
```

Prosite for DKFZphtes3_35p22.3

PS00004	136->140	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	310->314	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	348->352	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	61->64	PKC_PHOSPHO_SITE	PDOC00005
PS00005	73->76	PKC_PHOSPHO_SITE	PDOC00005
PS00005	90->93	PKC_PHOSPHO_SITE	PDOC00005
PS00005	152->155	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00005	282->285	PKC_PHOSPHO_SITE	PDOC00005
PS00005	315->318	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	446->449	PKC_PHOSPHO_SITE	PDOC00005
PS00006	61->65	CK2_PHOSPHO_SITE	PDOC00006
PS00006	460->464	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	511->515	CK2_PHOSPHO_SITE	PDOC00006
PS00007	93->100	TYR_PHOSPHO_SITE	PDOC00007
PS00007	92->100	TYR_PHOSPHO_SITE	PDOC00007
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	101->107	MYRISTYL	PDOC00008
PS00008	230->236	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	366->372	MYRISTYL	PDOC00008
PS00008	441->447	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_35p22.3)

DKFZphtes3_4b4
----

group: testes derived

DKFZphtes3_4b4 encodes a novel 497 amino acid protein similar to SCP proteins and a human trypsin inhibitor.

The novel protein contains an extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 2, predicted by Prosite and Pfam. This domain is found in a variety of extracellular proteins from eukaryotes that have been found to be evolutionary related. The exact function of these proteins is not yet known. In addition, the protein is similar to a human trypsin inhibitor.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes or as a new protease inhibitor.

strong similarity to trypsin inhibitor

might be a new protease inhibitor?

Sequenced by AGOWA

Locus: /map="333.4 cR from top of Chr16 linkage group"

Insert length: 4574 bp
Poly A stretch at pos. 4551, polyadenylation signal at pos. 4539

```
   1 GGCGGCTGCT CCCATTGAGC TGTCTGCTCG CTGTGCCCGC TGTGCCTGCT
  51 GTGCCCGCGC TGTCGCCGCT GCTACCGCGT CTGCTGGACG CGGGAGACGC
 101 CAGCGAGCTG GTGATTGGAG CCCTGCGGAG AGCTCAAGCG CCCAGCTCTG
 151 CCCGAGGAGC CCAGGCTGCC CCGTGAGTCC CATAGTTGCT GCAGGAGTGG
 201 AGCCATGAGC TGCGTCCTGG GTGGTGTCAT CCCCTTGGGG CTGCTGTTCC
 251 TGGTCTGCGG ATCCCAAGGC TACCTCCTGC CCAACGTCAC TCTCTTAGAG
 301 GAGCTGCTCA GCAAATACCA GCACAACGAG TCTCACTCCC GGGTCCGCAG
 351 AGCCATCCCC AGGGAGGACA AGGAGGAGAT CCTCATGCTG CACAACAAGC
 401 TTCGGGGCCA GGTGCAGCCT CAGGCCTCCA ACATGGAGTA CATGACCTGG
 451 GATGACGAAC TGGAGAAGTC TGCTGCAGCG TGGGCCAGTC AGTGCATCTG
 501 GGAGCACGGG CCCACCAGTC TGCTGGTGTC CATCGGGCAG AACCTGGGCG
 551 CTCACTGGGG CAGGTATCGC TCTCCGGGGT TCCATGTGCA GTCCTGGTAT
 601 GACGAGGTGA AGGACTACAC CTACCCCTAC CCGAGCGAGT GCAACCCCTG
 651 GTGTCCAGAG AGGTGCTCGG GGCCTATGTG CACGCACTAC ACACAGATAG
 701 TTTGGGCCAC CACCAACAAG ATCGGTTGTG CTGTGAACAC CTGCCGGAAG
 751 ATGACTGTCT GGGGAGAAGT TTGGGAGAAC GCGGTCTACT TTGTCTGCAA
 801 TTATTCTCCA AAGGGGAACT GGATTGGAGA AGCCCCCTAC AAGAATGGCC
 851 GGCCCTGCTC TGAGTGCCCA CCCAGCTATG GAGGCAGCTG CAGGAACAAC
 901 TTGTGTTACC GAGAAGAAAC CTACACTCCA AAACCTGAAA CGGACGAGAT
 951 GAATGAGGTG GAAACGGCTC CCATTCCTGA AGAAAACCAT GTTTGGCTCC
1001 AACCGAGGGT GATGAGACCC ACCAAGCCCA AGAAAACCTC TGCGGTCAAC
1051 TACATGACCC AAGTCGTCAG ATGTGACACC AAGATGAAGG ACAGGTGCAA
1101 AGGGTCCACG TGTAACAGGT ACCAGTGCCC AGCAGGCTGC CTGAACCACA
1151 AGGCGAAGAT CTTTGGAACT CTGTTCTATG AAAGCTCGTC TAGCATATGC
1201 CGCGCCGCCA TCCACTACGG GATCCTGGAT GACAAGGGAG GCCTGGTGGA
1251 TATCACCAGG AACGGGAAGG TCCCCTTCTT CGTGAAGTCT GAGAGACACG
1301 GCGTGCAGTC CCTCAGCAAA TACAAACCTT CCAGCTCATT CATGGTGTCA
1351 AAAGTGAAAG TGCAGGATTT GGACTGCTAC ACGACCGTTG CTCAGCTGTG
1401 CCCGTTTGAA AAGCCAGCAA CTCACTGCCC AAGAATCCAT TGTCCGGCAC
1451 ACTGCAAAGA CGAACCTTCC TACTGGGCTC CGGTGTTTGG AACCAACATC
1501 TATGCAGATA CCTCAAGCAT CTGCAAGACA GCCGTGCACG CGGGAGTCAT
1551 CAGCAACGAG AGTGGGGGTG ACGTGGACGT GATGCCCGTG GATAAAAAGA
1601 AGACCTACGT GGGCTCGCTC AGGAATGGAG TTCAGTCTGA AAGCCTGGGG
1651 ACTCCTCGGG ATGGAAAGGC CTTCCGGATC TTTGCTGTCA GGCAGTGAAT
1701 TTCCAGCACC AGGGGAGAAG GGGCGTCTTC AGGAGGGCTT CGGGGTTTTG
1751 CTTTTATTTT TATTTTGTCA TTGCGGGGTA TATGGAGAGT CAGGAAACTT
1801 CCTTTGACTG ATGTTCAGTG TCCATCACTT TGTGGCCTGT GGGTGAGGTG
1851 ACATCTCATC CCCTCACTGA AGCAACAGCA TCCCAAGGTG CTCAGCCGGA
1901 CTCCCTGGTG CCTGATCCTG CTGGGGCCCG GGGGTCTCCA TCTGGACGTC
1951 CTCTCTCCTT TAGAGATCTG AGCTGTCTCT TAAAGGGGAC AGTTGCCCAA
2001 AATGTTCCTT GCTATGTGTT CTTCTGTTGG TGGAGGAAGT TGATTTCAAC
2051 CTCCCTGCCA AAAGAACAAA CCATTTGAAG CTCACAATTG TGAAGCATTC
2101 ACGGCGTCGG AAGAGGCCTT TTGAGCAAGC GCCAATGAGT TTCAGGAATG
2151 AAGTAGAAGG TAGTTATTTA AAAATAAAAA ACACAGTCCG TCCCTACCAA
2201 TAGAGGAAAA TGGTTTTAAT GTTTGCTGGT CAGACAGACA AATGGGCTAG
2251 AGTAAGAGGG CTGCGGGTAT GAGAGACCCC GGCTCCGCCC TGGCACGTGT
2301 CCTTGCTGGC GGCCCGCCAC AGGCCCCCTT CAATGGCCGC ATTCAGGATG
2351 GCTCTATACA CAGCAGTGCT GGTTTATGTA GAGTTCAGCA GTCACTTCAG
2401 AGATGTATCT TGTCTTTGTC AGGCCCTTCA TCTTCATGGC CCACCTGTTT
2451 TCTGCCGTGA CCTTTGGTCC CATTGAGGAC TAAGGATCGG GACCCTTTCT
```

```
2501 TTACCCCCTA CCCATTGTGG CTCCCACCCT GCCTCGGACT GGTTTACGTG
2551 TCCTGGTTCA CACCCAGGAC TTTTCTTTGC AAGCGAACCT GTTTGAAGCC
2601 CAAGTCTTAA CTCCTGGTCT CGTAAGGTTC CACTGAGACG AGATGTCTGA
2651 GAACAACCAA AGAAGGCCTG CTCTTTGCTG CTTTTAAAAA ATGACAATTA
2701 AATGTGCAGA TTCCCCACGC ACCCGATGAC CTATTTTTTC AGCCGTGGGA
2751 GGAATGGAGT CTTTGGTACA TTCCTCACCG AGGTTAGCAG CTCAGTTTGT
2801 GGTTATGAAA CCGTCTGTGG CCTCATGACA GCGAGAGATG GGAATACACT
2851 AGAAGGATCT CTTTTCCTGT TTTCGTGAAA CGACTCTTGC CAAACGTTCC
2901 CGAGGCGCCA AGGAGTGTAG TACACCCTGG CTGCCATCAC TCTATAAAAG
2951 TGCTTCATGA GCCCAGACCA AAAGCCCACA GTGAAATGAA GTACCCTTTT
3001 GTAAATAGCA TTTTTTTGCA GAAGGTGAAA ATTCCACTCT CTACCACCGG
3051 GCCAGCCAAT AGATCACTTT GGTGAATGCT AGTTTCAAAT TTGATTCAAA
3101 ATATTTCTTA GGTGAAAGAA CTAGCAGAAA GTCAAAAACT AAGATACTGT
3151 AGACTGGACA AGAAATTCTA CCTGGGCACC TAGGTGATGC CTTCTTTCTT
3201 TGATTGCCTT TCTAATAAAT GCAGAATCTG AAGGTAAATA GGTTTAAAAC
3251 AAAACAAAAA CCCACCCCTT TAAGGAGTTG GTAAAAAGCA GTTCAACTCT
3301 TAGCTTGACT GAGCTAAAAT TCACAGGACT ACGTGCTTTG TGCATTGTAG
3351 TCTAGTCGTA ATTCATAGGT ACTGACTCCT CAGCCCCAAA TGTCGGAGAG
3401 GAAGAATTCG GTCAGCCTGT CAGGTCGTGA GTCCAGTTAC CACCAAACAT
3451 CTGGGAAACT TCTGGGTGCT GGGTGCTCTG CTGCTGGACT TTTGTGGCTG
3501 TGTCTGTGTC TGCAAGATAA ATTAGATCGC CCTGTGGGGT TTGCAGAATT
3551 AGTGAAGGGT CCAGGACGAT CCCAGTGGGC TCGCTTCCAA AGCATCCCAC
3601 TCAAGGGAGA CTTGAAACTT CCAGTGTGAG TTGACCCCAT CATTTAAAAA
3651 TAAAGTCCCC GGGTTCCTTA ATGCCTCCTT CACTGGGCCT TCCTAGCAGG
3701 ATAGAAAGTC CTTGCCCAGA GCAGGACCTG GCTGTCTTTT TTTTTTTTTT
3751 TTTCCCGAGA CCAAGTTTCA CTCTGTTGCC CAAGGTAGAG TGCAGTGGCG
3801 TGATCTCTGC TCATTGCAAC TGCCGCCTCC CGGGTTCAAG CAATTCTCAT
3851 GCATCAGCCT CCCAAGTACC TGGGACTACA GGCGTGAGCT ACCATGCCCG
3901 GCTAATTTTT GTATTTTTAG TAGAGATGGG GTTTCATTAT GTTGGCCAGG
3951 CTGGTCTCGA ACTCCTTACC TCAGGTGATC CACCCACCTT GGCCTCCCGA
4001 AGTGCTGGGA TTACAGGCAT GAGCCACTGC GCCCGGCCAT GGACCTGGCT
4051 GTCTTTATCA TCCCCACAAA CATTTGAAA CTGGAATATT TGTCTTCAGA
4101 AAATGGAAAC AAGACTATAA ATGATAAGCC CTGTCCCTAG CACCACCTCT
4151 CCTGTGTGTG GAATAGAGGC CCCTCGTGCT ACCAACACTT ACCCTGTGTT
4201 TAAAAAGATC TTGTACCAAG CCAACGGCGT TCCTGGCTCT CCTGCCCACA
4251 GGATGAACAT TTTCGGCTTC CTTAGGAGTT TTGCCCTACC GTATTCCAAA
4301 GCGTGTGCTG GTTTCTCATA TTGTCTGTAG GCTCACTCAG CCCGCAGTTT
4351 ATGTGTGTGC TTTTTTCTAT GAAAAATGAT GTATTTTGCT ACTTCCTGTG
4401 TACAAAGTTT TATTGTAAAT GTTTTTTGTG CTTTGCATGA ACAGGGGCCA
4451 CGTTGTTGCA ATTGTTTCAG TAGAACTGGT TTGATTTCTA AAATGTTCCT
4501 GTAACATATC TTTTATGAAC AAATCTGAAC AATTTGTGAA ATAAAACATT
4551 GAAAACCAAA AAAAAAAAAA AAAA
```

BLAST Results

Entry HS834352 from database EMBL:
human STS WI-15502.
Score = 1331, P = 5.4e-54, identities = 287/301

Medline entries

98146272:
cDNA cloning of a novel trypsin inhibitor with similarity to pathogenesis-related proteins, and its frequent expression in human brain cancer cells.

Peptide information for frame 1

ORF from 205 bp to 1695 bp; peptide length: 497
Category: strong similarity to known protein

```
  1 MSCVLGGVIP LGLLFLVCGS QGYLLPNVTL LEELLSKYQH NESHSRVRRA
 51 IPREDKEEIL MLHNKLRGQV QPQASNMEYM TWDDELEKSA AAWASQCIWE
101 HGPTSLLVSI GQNLGAHWGR YRSPGFHVQS WYDEVKDYTY PYPSECNPWC
151 PERCSGPMCT HYTQIVWATT NKIGCAVNTC RKMTVWGEVW ENAVYFVCNY
201 SPKGNWIGEA PYKNGRPCSE CPPSYGGSCR NNLCYREETY TPKPETDEMN
251 EVETAPIPEE NHVWLQPRVM RPTKPKKTSA VNYMTQVVRC DTKMKDRCKG
301 STCNRYQCPA GCLNHKAKIF GTLFYESSSS ICRAAIHYGI LDDKGGLVDI
351 TRNGKVPFFV KSERHGVQSL SKYKPSSSFM VSKVKVQDLD CYTTVAQLCP
401 FEKPATHCPR IHCPAHCKDE PSYWAPVFGT NIYADTSSIC KTAVHAGVIS
```

```
451 NESGGDVDVM PVDKKKTYVG SLRNGVQSES LGTPRDGKAF RIFAVRQ
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4b4, frame 1

TREMBLNEW:AF109674_1 gene: "Lgl1"; product: "late gestation lung protein 1"; Rattus norvegicus late gestation lung protein 1 (Lgl1) mRNA, complete cds., N = 1, Score = 968, P = 1.9e-97

TREMBL:D45027_1 product: "25 kDa trypsin inhibitor"; Homo sapiens mRNA for 25 kDa trypsin inhibitor, complete cds., N = 1, Score = 738, P = 4.5e-73

TREMBL:AB009609_1 gene: "HrTT-1"; Halocynthia roretzi HrTT-1 mRNA, complete cds., N = 1, Score = 345, P = 2e-31

PIR:JC5308 testis-specific, vespid, and pathogenesis-related protein 1 precursor - human, N = 1, Score = 337, P = 1.7e-30

```
>TREMBLNEW:AF109674_1 gene: "Lgl1"; product: "late gestation lung protein
     1"; Rattus norvegicus late gestation lung protein 1 (Lgl1) mRNA, complete
     cds.
            Length = 188

  HSPs:

 Score = 968 (145.2 bits), Expect = 1.9e-97, P = 1.9e-97
 Identities = 160/185 (86%), Positives = 170/185 (91%)

Query:    61 MLHNKLRGQVQPQASNMEYMTWDDELEKSAAAWASQCIWEHGPTSLLVSIGQNLGAHWGR 120
             MLHNKLRGQV P ASNMEYMTWD+ELE+SAAAWA +C+WEHGP SLLVSIGQNL  HWGR
Sbjct:     1 MLHNKLRGQVYPPASNMEYMTWDEELERSAAAWAQRCLWEHGPASLLVSIGQNLAVHWGR 60

Query:   121 YRSPGFHVQSWYDEVKDYTYPYPSECNPWCPERCSGPMCTHYTQIVWATTNKIGCAVNTC 180
             YRSPGFHVQSWYDEVKDYTYPYP ECNPWCPERCSG MCTHYTQ+VWATTNKIGCAV+TC
Sbjct:    61 YRSPGFHVQSWYDEVKDYTYPYPHECNPWCPERCSGAMCTHYTQMVWATTNKIGCAVHTC 120

Query:   181 RKMTVWGEVWENAVYFVCNYSPKGNWIGEAPYKNGRPCSECPPSYGGSCRNNLCYREETY 240
             R M+VWG++WENAVY VCNYSPKGNWIGEAPYK+GRPCSECP SYGG CRNNLCYREE Y
Sbjct:   121 RSMSVWGDIWENAVYLVCNYSPKGNWIGEAPYKHGRPCSECPSSYGGGCRNNLCYREEHY 180

Query:   241 TPKPE 245
              KPE
Sbjct:   181 HQKPE 185
```

Pedant information for DKFZphtes3_4b4, frame 1

Report for DKFZphtes3_4b4.1

```
[LENGTH]        497
[MW]            55920.00
[pI]            8.36
[HOMOL]         TREMBL:D45027_1 product: "25 kDa trypsin inhibitor"; Homo sapiens mRNA for 25
kDa trypsin inhibitor, complete cds. 6e-78
[FUNCAT]        99 unclassified proteins      [S. cerevisiae, YJL078c] 8e-12
[BLOCKS]        BL01009E Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS]        BL01009D Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS]        BL01009C Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[BLOCKS]        BL01009A Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins
[PIRKW]         glycoprotein 5e-22
[PIRKW]         blocked amino end 5e-13
[PIRKW]         brain 9e-30
[PIRKW]         hydrolase 4e-09
[PIRKW]         hemolymph coagulation 4e-09
[PIRKW]         zymogen 4e-09
[PIRKW]         alternative splicing 4e-09
[PIRKW]         sperm 5e-22
[PIRKW]         viroid-induced protein 2e-11
[PIRKW]         venom 6e-18
[PIRKW]         pyroglutamic acid 2e-11
[PIRKW]         transmembrane protein 2e-10
[PIRKW]         serine proteinase 4e-09
[SUPFAM]        C-type lectin homology 4e-09
[SUPFAM]        trypsin homology 4e-09
```

[SUPFAM]	complement factor H repeat homology 4e-09	
[SUPFAM]	cysteine-rich secretory protein 1 6e-24	
[SUPFAM]	pathogenesis-related leaf protein 7e-15	
[PROSITE]	MYRISTYL	8
[PROSITE]	CAMP_PHOSPHO_SITE	3
[PROSITE]	CK2_PHOSPHO_SITE	6
[PROSITE]	TYR_PHOSPHO_SITE	1
[PROSITE]	PKC_PHOSPHO_SITE	8
[PROSITE]	ASN_GLYCOSYLATION	3
[PROSITE]	SCP_AG5_PR1_SC7_2	1
[PFAM]	SCP-like extracellular Proteins	
[KW]	All_Beta	
[KW]	SIGNAL_PEPTIDE 23	
[KW]	LOW_COMPLEXITY	1.21 %

```
SEQ  MSCVLGGVIPLGLLFLVCGSQGYLLPNVTLLEELLSKYQHNESHSRVRRAIPREDKEEIL
SEG  .............................xxxxxx.........................
PRD  ccceeeeeceeeeeeeeccccccccccchhhhhhhhhhhhhccccchhhhhhhhcchhhhhh

SEQ  MLHNKLRGQVQPQASNMEYMTWDDELEKSAAAWASQCIWEHGPTSLLVSIGQNLGAHWGR
SEG  ............................................................
PRD  hhhhhhhcccccccccchhhhhhhhhhhhhhhhhhhhcccccccccccccccceeeeeecc

SEQ  YRSPGFHVQSWYDEVKDYTYPYPSECNPWCPERCSGPMCTHYTQIVWATTNKIGCAVNTC
SEG  ............................................................
PRD  cccchhhhhhhhhhhcccccccccccccccccccccccceeeeeeecccccceeeec

SEQ  RKMTVWGEVWENAVYFVCNYSPKGNWIGEAPYKNGRPCSECPPSYGGSCRNNLCYREETY
SEG  ............................................................
PRD  ccccccccccceeeeeeeccccccccccccccccccccccccccccccccccccccccccc

SEQ  TPKPETDEMNEVETAPIPEENHVWLQPRVMRPTKPKKTSAVNYMTQVVRCDTKMKDRCKG
SEG  ............................................................
PRD  ccccccccccccccccccceeeeecccccccccccccceeeeeeeeeeeeecccccccccc

SEQ  STCNRYQCPAGCLNHKAKIFGTLFYESSSSICRAAIHYGILDDKGGLVDITRNGKVPFFV
SEG  ............................................................
PRD  ccccccccccccccccceeeeeeeeeccccceeeeeccccccccccceeeeecccceeee

SEQ  KSERHGVQSLSKYKPSSSFMVSKVKVQDLDCYTTVAQLCPFEKPATHCPRIHCPAHCKDE
SEG  ............................................................
PRD  ecccceeeeeeeeccccceeeeeeeeeccccceeeeeeeecccccccccccccccccc

SEQ  PSYWAPVFGTNIYADTSSICKTAVHAGVISNESGGDVDVMPVDKKKTYVGSLRNGVQSES
SEG  ............................................................
PRD  ccceeeeeceeecccccceeeeeeeeeccccccccccccceeecccceeeeeeeccceeeee

SEQ  LGTPRDGKAFRIFAVRQ
SEG  .................
PRD  cccccccccceeeeccc
```

Prosite for DKFZphtes3_4b4.1

PS00001	27->31	ASN_GLYCOSYLATION	PDOC00001
PS00001	41->45	ASN_GLYCOSYLATION	PDOC00001
PS00001	451->455	ASN_GLYCOSYLATION	PDOC00001
PS00004	181->185	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	276->280	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	464->468	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	179->182	PKC_PHOSPHO_SITE	PDOC00005
PS00005	201->204	PKC_PHOSPHO_SITE	PDOC00005
PS00005	228->231	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005
PS00005	362->365	PKC_PHOSPHO_SITE	PDOC00005
PS00005	471->474	PKC_PHOSPHO_SITE	PDOC00005
PS00005	483->486	PKC_PHOSPHO_SITE	PDOC00005
PS00006	29->33	CK2_PHOSPHO_SITE	PDOC00006
PS00006	75->79	CK2_PHOSPHO_SITE	PDOC00006
PS00006	81->85	CK2_PHOSPHO_SITE	PDOC00006
PS00006	130->134	CK2_PHOSPHO_SITE	PDOC00006
PS00006	453->457	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00007	385->393	TYR_PHOSPHO_SITE	PDOC00007
PS00008	111->117	MYRISTYL	PDOC00008
PS00008	115->121	MYRISTYL	PDOC00008
PS00008	174->180	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

PS00008	227->233	MYRISTYL	PDOC00008
PS00008	300->306	MYRISTYL	PDOC00008
PS00008	447->453	MYRISTYL	PDOC00008
PS00008	470->476	MYRISTYL	PDOC00008
PS01010	195->207	SCP_AG5_PR1_SC7_2	PDOC00772

Pfam for DKFZphtes3_4b4.1

```
HMM_NAME     SCP-like extracellular Proteins

HMM             *PQDEQDEWLNkHNDFRQQVGRGLETRGNPGPQPPAsNMnPMVWNDELAt
                 P + ++E+L  HN +R QV             P ASNM  M+W+DEL +
Query        52  PREDKEEILMLHNKLRGQVQ------------PQASNMEYMTWDDELEK    88

HMM             IAQnWANQCiFDHHDCCWNHsnYPYGQNIAWWSsTANnPWnWssMIQMWY
                A  WA+QCI +H ++ +  S     GQN+ +  +    ++++    +Q+WY
Query        89 SAAAWASQCIWEHGPTSLLVSI---GQNLGAHWG---RYRSPGFHVQSWY    132

HMM             NEvkDYNYNWNTCkGG.....NNFmVCGHYTQMVWRnTfrIGCGRYICYC
                +EVKDY Y    +              + +C HYTQ+VW+ T +IGC+    C+
Query       133 DEVKDYTYPYPSECNPWCPERCSGPMCTHYTQIVWATTNKIGCAVNTCRK    182

HMM             NNNWrKPDPWKhkWYYVCNYCPpGNYmN*
                 + W    + W+   +Y VCNY P+GN+++
Query       183 MTVW--GEVWENAVYFVCNYSPKGNWIG    208
```

DKFZphtes3_4f17
----

group: testes derived

DKFZphtes3_4f17 encodes a novel 656 amino acid protein with weak similarity to methyl-CpG-binding proteins.

Methylation at the DNA sequence 5'-CpG is required for mammalian development. Methyl-CpG-binding proteins bind specifically to methylated DNA via a related amino acid motif and can repress transcription. The novel protein does not contain such a motife.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to methyl-CpG-binding protein

extension of HS557771/HSZ78337,
there are some differences to these sequences

Sequenced by AGOWA

Locus: /map="18"

Insert length: 2320 bp
Poly A stretch at pos. 2266, polyadenylation signal at pos. 2251

```
   1 GGCAGGTTCG CGGGTCGCTG GCGGGGGTCG TGAGGGAGTG CGCCGGGAGC
  51 GGAGATATGG AGGGAGATGG TTCAGACCCA GAGCCTCCAG ATGCCGGGGA
 101 GGACAGCAAG TCCGAGAATG GGGAGAATGC GCCCATCTAC TGCATCTGCC
 151 GCAAACCGGA CATCAACTGC TTCATGATCG GGTGTGACAA CTGCAATGAG
 201 TGGTTCCATG GGGACTGCAT CCGGATCACT GAGAAGATGG CCAAGGCCAT
 251 CCGGGAGTGG TACTGTCGGG AGTGCAGAGA GAAAGACCCC AAGCTAGAGA
 301 TTCGCTATCG GCACAAGAAG TCACGGGAGC GGGATGGCAA TGAGCGGGAC
 351 AGCAGTGAGC CCCGGGATGA GGGTGGAGGG CGCAAGAGGC CTGTCCCTGA
 401 TCCAGACCTG CAGCGCCGGG CAGGGTCAGG GACAGGGGTT GGGGCCATGC
 451 TTGCTCGGGG CTCTGCTTCG CCCCACAAAT CCTCTCCGCA GCCCTTGGTG
 501 GCCACACCCA GCCAGCATCA CCAGCAGCAG CAGCAGCAGA TCAAACGGTC
 551 AGCCCGCATG TGTGGTGAGT GTGAGGCATG TCGGCGCACT GAGGACTGTG
 601 GTCACTGTGA TTTCTGTCGG GACATGAAGA AGTTCGGGGG CCCCAACAAG
 651 ATCCGGCAGA AGTGCCGGCT GCGCCAGTGC CAGCTGCGGG CCCGGGAATC
 701 GTACAAGTAC TTCCCTTCCT CGCTCTCACC AGTGACGCCC TCAGAGTCCC
 751 TGCCAAGGCC CCGCCGGCCA CTGCCCACCC AACAGCAGCC ACAGCCATCA
 801 CAGAAGTTAG GGCGCATCCG TGAAGATGAG GGGGCAGTGG CGTCATCAAC
 851 AGTCAAGGAG CCTCCTGAGG CTACAGCCAC ACCTGAGCCA CTCTCAGATG
 901 AGGACCTACC TCTGGATCCT GACCTGTATC AGGACTTCTG TGCAGGGGCC
 951 TTTGATGACC ATGGCCTGCC CTGGATGAGC GACACAGAAG AGTCCCCATT
1001 CCTGGACCCC GCGCTGCGGA AGAGGGCAGT GAAAGTGAAG CATGTGAAGC
1051 GTCGGGAGAA GAAGTCTGAG AAGAAGAAGG AGGAGCGATA CAAGCGGCAT
1101 CGGCAGAAGC AGAAGCACAA GGATAAATGG AAACACCCAG AGAGGGCTGA
1151 TGCCAAGGAC CCTGCGTCAC TGCCCCAGTG CCTGGGGCCC GGCTGTGTGC
1201 GCCCCGCCCA GCCCAGCTCC AAGTATTGCT CAGATGACTG TGGCATGAAG
1251 CTGGCAGCCA ACCGCATCTA CGAGATCCTC CCCCAGCGCA TCCAGCAGTG
1301 GCAGCAGAGC CCTTGCATTG CTGAAGAGCA CGGCAAGAAG CTGCTCGAAC
1351 GCATTCGCCG AGAGCAGCAG AGTGCCCGCA CCCGCCTTCA GGAAATGGAA
1401 CGCCGATTCC ATGAGCTTGA GGCCATCATT CTACGTGCCA AGCAGCAGGC
1451 TGTGCGCGAG GATGAGGAGA GCAACGAGGG TGACAGTGAT GACACAGACC
1501 TGCAGATCTT CTGTGTTTCC TGTGGGCACC CCATCAACCC ACGTGTTGCC
1551 TTGCGCCACA TGGAGCGCTG CTACGCCAAG TATGAGAGCC AGACGTCCTT
1601 TGGGTCCATG TACCCCACAC GCATTGAAGG GGCCACACGA CTCTTCTGTG
1651 ATGTGTATAA TCCTCAGAGC AAAACATACT GTAAGCGGCT CCAGGTGCTG
1701 TGCCCCGAGC ACTCACGGGA CCCCAAAGTG CCAGCTGACG AGGTATGCGG
1751 GTGCCCCCTT GTACGTGATG TCTTTGAGCT CACGGGTGAC TTCTGCCGCC
1801 TGCCCAAGCG CCAGTGCAAT CGCCATTACT GCTGGGAGAA GCTGCGGCGT
1851 GCGGAAGTGG ACTTGGAGCG CGTGCGTGTG TGGTACAAGC TGGACGAGCT
1901 GTTTGAGCAG GAGCGCAATG TGCGCACAGC CATGACAAAC CGCGCGGGAT
1951 TGCTGGCCCT GATGCTGCAC CAGACGATCC AGCACGATCC CCTCACTACC
2001 GACCTGCGCT CCAGTGCCGA CCGCTGAGCC TCCTGGCCCG GACCCCTTAC
2051 ACCCTGCATT CCAGATGGGG GAGCCGCCCG GTGCCCGTGT GTCCGTTCCT
2101 CCACTCATCT GTTTCTCCGG TTCTCCCTGT GCCCATCCAC CGGTTGACCG
2151 CCCATCTGCC TTTATCAGAG GGACTGTCCC CGTCGACATG TTCAGTGCCT
2201 GGTGGGGCTG CGGAGTCCAC TCATCCTTGC CTCCTCTCCC TGGGTTTTGT
2251 TAATAAAATT TTGAAGAAAC CAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2301 AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

Entry HS557771 from database EMBLEST:
Human chromosome 18 clone 2 mRNA sequence.
Score = 7582, P = 0.0e+00, identities = 1560/1598

Entry HSZ78337 from database EMBLEST:
H.sapiens mRNA, expressed sequence tag ICRFp507H02194 (5')
Score = 6339, P = 9.0e-281, identities = 1307/1347

Entry HS095149 from database EMBL:
human STS WI-6941.
Score = 1210, P = 2.2e-49, identities = 246/251

Medline entries

98449942:
Identification and characterization of a family of mammalian methyl-CpG binding proteins.

9824997:
Gene silencing by methyl-CpG-binding proteins.

Peptide information for frame 3

ORF from 57 bp to 2024 bp; peptide length: 656
Category: similarity to known protein

```
  1 MEGDGSDPEP PDAGEDSKSE NGENAPIYCI CRKPDINCFM IGCDNCNEWF
 51 HGDCIRITEK MAKAIREWYC RECREKDPKL EIRYRHKKSR ERDGNERDSS
101 EPRDEGGGRK RPVPDPDLQR RAGSGTGVGA MLARGSASPH KSSPQPLVAT
151 PSQHHQQQQQ QIKRSARMCG ECEACRRTED CGHCDFCRDM KKFGGPNKIR
201 QKCRLRQCQL RARESYKYFP SSLSPVTPSE SLPRPRRPLP TQQQPQPSQK
251 LGRIREDEGA VASSTVKEPP EATATPEPLS DEDLPLDPDL YQDFCAGAFD
301 DHGLPWMSDT EESPFLDPAL RKRAVKVKHV KRREKKSEKK KEERYKRHRQ
351 KQKHKDKWKH PERADAKDPA SLPQCLGPGC VRPAQPSSKY CSDDCGMKLA
401 ANRIYEILPQ RIQQWQQSPC IAEEHGKKLL ERIRREQQSA RTRLQEMERR
451 FHELEAIILR AKQQAVREDE ESNEGDSDDT DLQIFCVSCG HPINPRVALR
501 HMERCYAKYE SQTSFGSMYP TRIEGATRLF CDVYNPQSKT YCKRLQVLCP
551 EHSRDPKVPA DEVCGCPLVR DVFELTGDFC RLPKRQCNRH YCWEKLRRAE
601 VDLERVRVWY KLDELFEQER NVRTAMTNRA GLLALMLHQT IQHDPLTTDL
651 RSSADR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4f17, frame 3

TREMBL:CEF52B11_4 gene: "F52B11.1"; Caenorhabditis elegans cosmid F52B11, N = 2, Score = 316, P = 8.8e-27

TREMBL:HSAB2331_1 gene: "KIAA0333"; Human mRNA for KIAA0333 gene, partial cds., N = 2, Score = 163, P = 2.8e-13

TREMBL:SPCC594_5 gene: "SPCC594.05c"; product: "putative transcriptional regulatory protein, phd finger containing"; S.pombe chromosome III cosmid c594., N = 3, Score = 168, P = 3.6e-12

TREMBL:AF072240_1 gene: "Mbd1"; product: "methyl-CpG binding protein MBD1"; Mus musculus methyl-CpG binding protein MBD1 (Mbd1) mRNA, complete cds., N = 2, Score = 189, P = 7.6e-11

>TREMBL:CEF52B11_4 gene: "F52B11.1"; Caenorhabditis elegans cosmid F52B11
Length = 523

HSPs:

Score = 316 (47.4 bits), Expect = 8.8e-27, Sum P(2) = 8.8e-27
Identities = 100/336 (29%), Positives = 167/336 (49%)

```
Query:  333 REKKSEKKKEERYKRHRQ-KQKHKDKWKHPERADAKDPASLP-QCLGPGCVRPAQPSSKY 390
            +++K+    E  Y  +R  +Q+  D +    + +A +P    P QCL P C+  ++  SKY
Sbjct:  118 QQRKANIINERDYVPNRPTRQQSADLRRKRTQLNA-EPDKHPRQCLNPNCIYESRIDSKY 176

Query:  391 CSDDCGMKLAANRIYEILPQRIQQW-----QQSPCIAEEHGKKLLERIRREQQSARTRLQ 445
            CSD+CG +LA  R+ EILP R +Q+            P    E+  K    +I RE Q    +
Sbjct:  177 CSDECGKELARMRLTEILPNRCKQYFFEGPSGGPRSLEDEIKPKRAKINREVQKLTESEK 236

Query:  446 EMERRFHEL-EAIILRAKQQAVREDEESNEGDSDDTDLQIFCVSCGHPINPRVAL-RHME 503
             M    ++L E I  + K Q +  +E          D +L   C+ CG P  P +   +H+E
Sbjct:  237 NMMAFLNKLVEFIKTQLKLQPLGTEERY------DDNLYEGCIVCGLPDIPLLKYTKHIE 290

Query:  504 RCYAKYESQTSFGSMYPTRIEGATRLFCDVYNPQSKTYCKRLQVLCPEHSRDPKVPADEV 563
             C+A+ E    SFG+   P +         +C+ Y+ ++ ++CKRL+ LCPEH +        +V
Sbjct:  291 LCWARSEKAISFGA--PEK--NNDMFYCEKYDSRTNSFCKRLKSLCPEHRKLGDEQHLKV 346

Query:  564 CGCP------------LVRDVFELTGDF----CRLPKRQCNRHYCWEKLRRAEVDLERVR 607
            CG P             V ++ E+    F    CR  K  C++H+ W     R  ++LE+
Sbjct:  347 CGYPKKWEDGMIETAKTVSELIEMEDPFGEEGCRTKKDACHKHHKWIPSLRGTIELEQAC 406

Query:  608 VWYKLDELFEQ--ERNVRTAMTNRAGLLALMLHQTIQHDPLTTDLRSSA 654
            ++ K+ EL   +   + N      T  A  L++M+H+     +  +  LR+ A
Sbjct:  407 LFQKMYELCHEMHKLNAHAEWTTNA--LSIMMHKQPSTEKCSFFLRNFA 453

 Score = 53 (8.0 bits), Expect = 8.8e-27, Sum P(2) = 8.8e-27
 Identities = 24/100 (24%), Positives = 41/100 (41%)

Query:  169 CGECEACRRTEDCGHCDFCR-----DMKK-FGGPNKIRQKCRLRQCQLRARESYKYFPSS 222
            C  C  C   ++CG C  CR      DM+K F        +K + RQ     + +     +
Sbjct:   17 CMNCIRCNDEKNCGTCWPCRNGKTCDMRKCFSAKRLYNEKVK-RQTDENLK-AIMAKTAQ 74

Query:  223 LSPVTPSESLPRPRRPLPTQQQPQPSQKLGRIR-EDEGAVASS 264
                  + +   P  P+  +QQ +  +K GR +    G  A++
Sbjct:   75 REAAHQAATTAPSAPVVIEQQVE-KKKRGRKKGSGNGGAAAA 116

 Score = 48 (7.2 bits), Expect = 2.9e-26, Sum P(2) = 2.9e-26
 Identities = 13/39 (33%), Positives = 19/39 (48%)

Query:  179 EDCGHCDFCRDMKKFGG--PNKIRQKCRLRQCQLRARESY 216
            E C +C  C D K  G   P +  + C +R+C   A+  Y
Sbjct:   15 ERCMNCIRCNDEKNCGTCWPCRNGKTCDMRKC-FSAKRLY 53
```

Pedant information for DKFZphtes3_4f17, frame 3

Report for DKFZphtes3_4f17.3

[LENGTH]	656
[MW]	75711.71
[pI]	8.61
[HOMOL]	TREMBL:CEF52B11_4 gene: "F52B11.1"; Caenorhabditis elegans cosmid F52B11 3e-25
[FUNCAT]	99 unclassified proteins [S. cerevisiae, YPL138c] 3e-10
[FUNCAT]	04.05.01.04 transcriptional control [S. cerevisiae, YNL097c] 2e-04
[PROSITE]	MYRISTYL 6
[PROSITE]	AMIDATION 2
[PROSITE]	CK2_PHOSPHO_SITE 8
[PROSITE]	TYR_PHOSPHO_SITE 3
[PROSITE]	GLYCOSAMINOGLYCAN 1
[PROSITE]	PKC_PHOSPHO_SITE 9
[KW]	All_Alpha
[KW]	LOW_COMPLEXITY 18.75 %
[KW]	COILED_COIL 4.57 %

```
SEQ    MEGDGSDPEPPDAGEDSKSENGENAPIYCICRKPDINCFMIGCDNCNEWFHGDCIRITEK
SEG    ............................................................
PRD    cccccccccccccccccccccccccceeeeeeccccceeeeeccccccccccccchhhhhh
COILS  ............................................................

SEQ    MAKAIREWYCRECREKDPKLEIRYRHKKSRERDGNERDSSEPRDEGGGRKRPVPDPDLQR
SEG    ............................................................
PRD    hhhhhhhhhhhcccccccccchhhhhhhhhcccccccccccccccccccccccccccccc
COILS  ............................................................

SEQ    RAGSGTGVGAMLARGSASPHKSSPQPLVATPSQHHQQQQQQIKRSARMCGECEACRRTED
SEG    ..................................xxxxxxxxx.................
PRD    ccccccccceeeccccccccccccccccccccchhhhhhhhhhhhhhhhhhcccccccccccc
COILS  ............................................................
```

```
SEQ    CGHCDFCRDMKKFGGPNKIRQKCRLRQCQLRARESYKYFPSSLSPVTPSESLPRPRRPLP
SEG    ...................xxxxxxxxxxxxxx.............xxxxxxxxxxxxxx
PRD    cccccccccccccccccchhhhhhhhhhhhhhhhhhcccccccccccccccccccccccc
COILS  ............................................................

SEQ    TQQQPQPSQKLGRIREDEGAVASSTVKEPPEATATPEPLSDEDLPLDPDLYQDFCAGAFD
SEG    xxxxxxxxx..................xxxxxxxxxxxxxxxxxxxxxxx..........
PRD    cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS  ............................................................

SEQ    DHGLPWMSDTEESPFLDPALRKRAVKVKHVKRREKKSEKKKEERYKRHRQKQKHKDKWKH
SEG    ....................xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD    ccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccchh
COILS  ............................................................

SEQ    PERADAKDPASLPQCLGPGCVRPAQPSSKYCSDDCGMKLAANRIYEILPQRIQQWQQSPC
SEG    ............................................................
PRD    hhhhcccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhccch
COILS  ............................................................

SEQ    IAEEHGKKLLERIRREQQSARTRLQEMERRFHELEAIILRAKQQAVREDEESNEGDSDDT
SEG    ...............................................xxxxxxxxxxxxx
PRD    hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccc
COILS  .................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.............

SEQ    DLQIFCVSCGHPINPRVALRHMERCYAKYESQTSFGSMYPTRIEGATRLFCDVYNPQSKT
SEG    x...........................................................
PRD    ceeeeeeeccccccccchhhhhhhhhhhhhhhcccccccccccccccceeeeeeccccccc
COILS  ............................................................

SEQ    YCKRLQVLCPEHSRDPKVPADEVCGCPLVRDVFELTGDFCRLPKRQCNRHYCWEKLRRAE
SEG    ............................................................
PRD    cchhhhhhcccccccccccceeeecccchhhhcccccccccccccccchhhhhhhhhhhh
COILS  ............................................................

SEQ    VDLERVRVWYKLDELFEQERNVRTAMTNRAGLLALMLHQTIQHDPLTTDLRSSADR
SEG    ........................................................
PRD    hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccc
COILS  ........................................................
```

Prosite for DKFZphtes3_4f17.3

PS00002	124->128	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00005	165->168	PKC_PHOSPHO_SITE	PDOC00005
PS00005	215->218	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	265->268	PKC_PHOSPHO_SITE	PDOC00005
PS00005	337->340	PKC_PHOSPHO_SITE	PDOC00005
PS00005	387->390	PKC_PHOSPHO_SITE	PDOC00005
PS00005	439->442	PKC_PHOSPHO_SITE	PDOC00005
PS00005	627->630	PKC_PHOSPHO_SITE	PDOC00005
PS00006	6->10	CK2_PHOSPHO_SITE	PDOC00006
PS00006	17->21	CK2_PHOSPHO_SITE	PDOC00006
PS00006	227->231	CK2_PHOSPHO_SITE	PDOC00006
PS00006	265->269	CK2_PHOSPHO_SITE	PDOC00006
PS00006	280->284	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	521->525	CK2_PHOSPHO_SITE	PDOC00006
PS00006	652->656	CK2_PHOSPHO_SITE	PDOC00006
PS00007	339->346	TYR_PHOSPHO_SITE	PDOC00007
PS00007	500->507	TYR_PHOSPHO_SITE	PDOC00007
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00008	42->48	MYRISTYL	PDOC00008
PS00008	123->129	MYRISTYL	PDOC00008
PS00008	125->131	MYRISTYL	PDOC00008
PS00008	129->135	MYRISTYL	PDOC00008
PS00008	259->265	MYRISTYL	PDOC00008
PS00008	396->402	MYRISTYL	PDOC00008
PS00009	107->111	AMIDATION	PDOC00009
PS00009	425->429	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_4f17.3)

DKFZphtes3_4f5
----

group: signal transduction

DKFZphtes3_4f5.3encodes a novel 790 amino acid protein similar to beta-transducins.

The protein contains 3 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a Cytochrome C family heme-binding site signature is present. The protein is larger (790 amino acids)than the usual eukaryotic G-beta transducins (about 340 amino acids).

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to S.pombe "beta-transducin"

complete cDNA, EST hits
complete cds,
on genomic level encoded by HS313D11, at least 7 exons these exons match
only partialy with the predicted transcripts in HS313D11

Sequenced by AGOWA

Locus: /map="16p13.3"

Insert length: 3166 bp
No poly A stretch found, no polyadenylation signal found

```
   1 GGCGGCTTCC GGCGCGGCGG TTCCGGACAA CCGTGCGCTT TTAGTAAAAG
  51 ATTGGGGTTC GCGCGGGGGA GAAGGGCTGC CCCGGGCCCT CTGGTTCTCG
 101 TCCCGCAGCG TCCGCTCCCC CGCGCCACTG CGCCGCTCCC AGGAACCCTG
 151 TACTCCGGGG TCGCCGGCTT CTCTCCTGCC TCCGGTCCCG CCAGACACCT
 201 CGAGCTCCTT AAGTAGCTCG GTCCTTGACG TCCCTCTGGG CCCTTCCCGC
 251 GTCTATCGCC TGAGTCCCCG GGCCCCTCTA GCCCTCTGTT CCCTCCCCTC
 301 TTTTGTTCCT CCCTAGAGCC CCGCCGCCCT CAGGGCTGAC AGTGTGGACG
 351 GCGGGAGTCT CCTCGCTCCC CTGCTGGGAT TGACTGACCG AGCGTTTAGT
 401 GACTGCCCAG ATCTGGCTGA TGGGGGTACC GAGAGGTGGC CTGGGCCGGG
 451 AATGTCCAGC TAGAGTCTTC CGTGGAAGTC AGACATGAAA CTGACAGGCC
 501 TAAGGGAAGC TAGGAAGTCC CCTCACCGCT CAGCCAGGGT GATGGGCTGG
 551 ACTGACAGAC TCCAGTGAAT TTGAGCTTGC CTGTCAGGCT GATTGGCTGA
 601 TAGACAGCCC TGGATTGGCT CACTAAGACT GACCAGCCCG GGACCAAGCA
 651 GTTCTGGGGT CCCAACCTGG GTGGAAGGTC TGAACTGATG ACCCACCCAG
 701 GCTGACCAGG CCAGCCCACC TCACTGACCT CCTGACCCCT GACCTCATCA
 751 CCTGTGCAGC CATGGAGAAG ATGTCCCGTG TGACCACAGC CCTGGGTGGC
 801 AGCGTGCTGA CAGGCCGCAC CATGCACTGC CACCTGGATG CTCCCGCCAA
 851 TGCCATCAGT GTGTGCCGCG ACGCAGCCCA GGTGGTCGTG GCAGGCCGTA
 901 GCATCTTCAA GATCTATGCC ATCGAGGAGG AACAGTTCGT GGAAAAGCTG
 951 AACCTGCGTG TGGGGCGCAA GCCTTCGCTT AACCTGAGCT GTGCTGACGT
1001 GGTCTGGCAC CAGATGGATG AGAACCTGCT GGCCACAGCA GCCACCAATG
1051 GCGTGGTGGT CACGTGGAAC CTGGGCCGGC CATCCCGCAA CAAGCAGGAC
1101 CAGCTGTTCA CAGAACACAA GCGCACGGTA AACAAAGTCT GCTTCCACCC
1151 CACCGAAGCC CACGTGCTGC TCAGTGGCTC CCAGGATGGC TTCATGAAGT
1201 GCTTTGACCT CCGCAGAAAG GACTCTGTCA GCACCTTCTC GGGCCAGTCG
1251 GAGAGCGTGC GGGACGTGCA GTTCAGTATC CGGGACTACT TCACCTTCGC
1301 CTCCACCTTT GAGAACGGCA ATGTGCAGCT CTGGGACATC CGGCGTCCCG
1351 ACCGGTGCGA GAGGATGTTC ACAGCCCACA ACGGACCCGT CTTCTGCTGC
1401 GACTGGCACC CCGAGGACAG GGGCTGGTTG GCCACTGGAG GGCGCGACAA
1451 GATGGTGAAG GTCTGGGACA TGACCACGCA CCGTGCCAAG GAGATGCACT
1501 GTGTGCAGAC CATCGCCTCG GTGGCCCGTG TGAAGTGGCG GCCAGAGTGC
1551 CGCCACCACC TGGCCACGTG CTCCATGATG GTGGACCACA ACATCTATGT
1601 TTGGGACGTG CGCCGGCCCT TCGTGCCAGC TGCCATGTTT GAGGAACACC
1651 GAGACGTCAC CACGGGAATT GCCTGGCGCC ACCCCCACGA CCCCTCCTTC
1701 CTGCTGTCTG GCTCCAAGGA CAGCTCGCTG TGCCAGCACC TGTTCCGCGA
1751 CGCCAGCCAG CCCGTCGAGC GCGCCAACCC TGAGGGCCTC TGCTACGGCC
1801 TCTTCGGGGA CCTGGCCTTC GCCGCCAAGG AGAGCCTCGT GGCTGCCGAG
1851 TCGGGGCGCA AGCCCTACAC TGGCGACCGG CGCCACCCCA TCTTCTTTAA
1901 GCGCAAGCTG GACCCTGCCG AGCCCTTCGC AGGCCTCGCC TCCAGTGCCC
1951 TCAGTGTCTT TGAGACGGAG CCAGGTGGCG GCGGCATGCG CTGGTTTGTG
2001 GACACAGCTG AGCGTTATGC GCTGGCTGGC CGGCCACTGG CCGAGCTCTG
2051 TGACCACAAC GCAAAGGTGG CTCGAGAGCT TGGCCGCAAC CAGGTGGCGC
2101 AAACGTGGAC CATGCTGCGG ATCATCTACT GCAGCCCTGG CCTAGTGCCC
2151 ACTGCAAACC TCAACCACAG TGTGGGCAAG GGTGGCTCCT GTGGCCTCCC
2201 GCTCATGAAC AGTTTCAACC TGAAGGATAT GGCCCCAGGG TTGGGCAGTG
2251 AGACGCGGCT GGACCGCAGC AAAGGAGATG CACGGAGCGA CACAGTTCTG
2301 CTCGACTCCT CGGCCACACT CATCACCAAT GAGGATAACG AGGAAACCGA
2351 GGGCAGCGAC GTACCTGCCG ACTACCTGCT GGGTGACGTG GAAGGTGAGG
```

```
2401 AGGACGAGCT GTACCTGCTG GATCCGGAAC ACGCGCACCC CGAGGACCCT
2451 GAGTGCGTGC TGCCGCAGGA GGCCTTTCCG CTGCGCCACG AGATCGTGGA
2501 CACGCCTCCC GGACCCGAGC ACCTGCAGGA CAAGGCCGAC TCCCCGCACG
2551 TGAGCGGCAG CGAGGCGGAT GTGGCCTCCC TGGCCCCCGT GGACTCCTCC
2601 TTCTCGCTCC TGTCTGTCTC ACACGCGCTC TACGACAGCC GCCTGCCGCC
2651 CGACTTCTTC GGCGTGCTGG TGCGCGACAT GCTGCACTTC TACGCTGAGC
2701 AGGGCGACGT GCAGATGGCT GTGTCTGTGC TCATCGTCCT GGGTGAACGG
2751 GTGCGCAAGG ACATCGACGA GCAGACCCAG GAGCACTGGT ACACTTCCTA
2801 CATCGACCTG CTGCAGCGCT TCCGCCTCTG GAACGTGTCC AACGAGGTGG
2851 TCAAGCTGAG CACCAGCCGC GCCGTCAGCT GCCTCAACCA GGCCTCCACC
2901 ACCCTGCACG TCAACTGCAG CCACTGCAAG CGGCCCATGA GCAGCCGGGG
2951 CTGGGTCTGC GACAGGTGCC ACCGCTGCGC CAGCATGTGT GCCGTCTGCC
3001 ACCACGTAGT CAAGGGTCTC TTCGTGTGGT GCCAGGGCTG CAGCCACGGC
3051 GGCCACCTGC AGCACATCAT GAAGTGGCTG GAAGGCAGCT CCCACTGTCC
3101 CGCAGGCTGC GGCCACCTCT GCGAGTACTC CTGACGGGGC ATCTGCTGGG
3151 CTTGCCCGGG CGGCCG
```

BLAST Results

Entry HS313D11 from database EMBL:
Human DNA sequence from cosmid 313D11 from a contig on the short arm of chromosome 16. Contains ESTs, STS and CpG islands.
Score = 6238, P = 0.0e+00, identities = 1318/1391

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 762 bp to 3131 bp; peptide length: 790
Category: similarity to known protein

```
  1 MEKMSRVTTA LGGSVLTGRT MHCHLDAPAN AISVCRDAAQ VVVAGRSIFK
 51 IYAIEEEQFV EKLNLRVGRK PSLNLSCADV VWHQMDENLL ATAATNGVVV
101 TWNLGRPSRN KQDQLFTEHK RTVNKVCFHP TEAHVLLSGS QDGFMKCFDL
151 RRKDSVSTFS GQSESVRDVQ FSIRDYFTFA STFENGNVQL WDIRRPDRCE
201 RMFTAHNGPV FCCDWHPEDR GWLATGGRDK MVKVWDMTTH RAKEMHCVQT
251 IASVARVKWR PECRHHLATC SMMVDHNIYV WDVRRPFVPA AMFEEHRDVT
301 TGIAWRHPHD PSFLLSGSKD SSLCQHLFRD ASQPVERANP EGLCYGLFGD
351 LAFAAKESLV AAESGRKPYT GDRRHPIFFK RKLDPAEPFA GLASSALSVF
401 ETEPGGGGMR WFVDTAERYA LAGRPLAELC DHNAKVAREL GRNQVAQTWT
451 MLRIIYCSPG LVPTANLNHS VGKGGSCGLP LMNSFNLKDM APGLGSETRL
501 DRSKGDARSD TVLLDSSATL ITNEDNEETE GSDVPADYLL GDVEGEEDEL
551 YLLDPEHAHP EDPECVLPQE AFPLRHEIVD TPPGPEHLQD KADSPHVSGS
601 EADVASLAPV DSSFSLLSVS HALYDSRLPP DFFGVLVRDM LHFYAEQGDV
651 QMAVSVLIVL GERVRKDIDE QTQEHWYTSY IDLLQRFRLW NVSNEVVKLS
701 TSRAVSCLNQ ASTTLHVNCS HCKRPMSSRG WVCDRCHRCA SMCAVCHHVV
751 KGLFVWCQGC SHGGHLQHIM KWLEGSSHCP AGCGHLCEYS
```

BLASTP hits

Entry YDSB_SCHPO from database SWISSPROT:
HYPOTHETICAL 93.2 KD TRP-ASP REPEATS CONTAINING PROTEIN C4F8.11 IN CHROMOSOME I. >TREMBL:SPAC4F8_11 gene: "SPAC4F8.11"; product: "beta-transducin"; S.pombe chromosome I cosmid c4F8.
Score = 404, P = 3.0e-42, identities = 169/639, positives = 278/639

Entry PEX7_HUMAN from database SWISSPROT:
PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7). >TREMBL:HSU76560_1 gene: "Pex7"; product: "peroxisome targeting signal 2 receptor"; Human peroxisome targeting signal 2 receptor (Pex7) mRNA, complete cds. >TREMBL:HSU88871_1 gene: "HsPEX7"; product: "HsPex7p"; Human HsPex7p (HsPEX7) mRNA, complete cds.
Score = 220, P = 1.1e-15, identities = 62/244, positives = 107/244

Entry PEX7_MOUSE from database SWISSPROT:
PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7). >TREMBL:MMU69171_1 product: "peroxisomal PTS2 receptor"; Mus musculus peroxisomal PTS2 receptor mRNA, complete cds.
Score = 214, P = 5.3e-15, identities = 60/240, positives = 106/240

Entry ATAC2294_7 from database TREMBL:
gene: "F11P17.7"; Arabidopsis thaliana chromosome I BAC F11P17 genomic sequence, complete sequence.
Score = 232, P = 3.4e-14, identities = 68/260, positives = 120/260

Entry S66835 from database PIR:
probable membrane protein YOL138c - yeast (Saccharomyces cerevisiae)
>TREMBL:SCYOL138C_1 S.cerevisiae chromosome XV reading frame ORF YOL138c
Score = 136, P = 2.5e-13, identities = 24/77, positives = 44/77

Alert BLASTP hits for DKFZphtes3_4f5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_4f5, frame 3

Report for DKFZphtes3_4f5.3

[LENGTH] 790
[MW] 88207.10
[pI] 6.05
[HOMOL] SWISSPROT:YDSB_SCHPO HYPOTHETICAL 93.2 KD TRP-ASP REPEATS CONTAINING PROTEIN C4F8.11 IN CHROMOSOME I. 9e-44
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YOL138c] 5e-16
[FUNCAT] 10.04.09 regulation of g-protein activity [S. cerevisiae, YBR195c] 3e-11
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YBR195c] 3e-11
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YBR195c] 3e-11
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YBR195c] 3e-11
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YBR195c] 3e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YCR072c beta-transducin family] 3e-10
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 9e-09
[FUNCAT] 04.01.04 rrna processing [S. cerevisiae, YLL011w] 1e-07
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL195w] 2e-07
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL195w] 2e-07
[FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR142c] 4e-07
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR142c] 4e-07
[FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YDR142c] 4e-07
[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YER107c] 4e-07
[FUNCAT] 04.07 rna transport [S. cerevisiae, YER107c] 4e-07
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER107c] 4e-07
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL003c] 5e-07
[FUNCAT] 06.13 proteolysis [S. cerevisiae, YGL003c] 5e-07
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 8e-07
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-06
[FUNCAT] 03.13 meiosis [S. cerevisiae, YLR129w] 3e-06
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 1e-05
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YCR057c] 1e-05
[FUNCAT] 06.07 protein modification (glycolsylation, acylation, myristylation, palmitylation, farnesylation and processing) [S. cerevisiae, YEL056w] 2e-04
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YOR272w] 6e-04
[SCOP] d1gotb_ 2.46.3.1.1 beta1-subunit of the signal-transducing 5e-06
[PIRKW] duplication 7e-10
[PIRKW] signal transduction 7e-08
[PIRKW] peroxisome 9e-06
[PIRKW] heterotrimer 7e-08
[PIRKW] GTP binding 7e-08
[PIRKW] peroxisome biogenesis 9e-06
[PIRKW] transmembrane protein 1e-14
[SUPFAM] MSI1 protein 7e-10
[SUPFAM] WD repeat homology 1e-14
[SUPFAM] GTP-binding regulatory protein beta chain 7e-08
[SUPFAM] PRL1 protein 3e-08
[SUPFAM] coatomer complex beta' chain 1e-06
[PROSITE] CYTOCHROME_C 1
[PROSITE] WD_REPEATS 3
[PROSITE] MYRISTYL 10
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 11

```
[PROSITE]     TYR_PHOSPHO_SITE     1
[PROSITE]     PKC_PHOSPHO_SITE     7
[PROSITE]     ASN_GLYCOSYLATION    4
[PFAM]        WD domain, G-beta repeats
[KW]          All_Beta
[KW]          3D
[KW]          LOW_COMPLEXITY    2.28 %
```

```
SEQ    MEKMSRVTTALGGSVLTGRTMHCHLDAPANAISVCRDAAQVVVAGRSIFKIYAIEEEQFV
SEG    ............................................................
1gotB  ............................................................

SEQ    EKLNLRVGRKPSLNLSCADVVWHQMDENLLATAATNGVVVTWNLGRPSRNKQDQLFTEHK
SEG    ............................................................
1gotB  .........................TTCEEEEEETTTEEEEEET-TTTCEEE--EEECCC

SEQ    RTVNKVCFHPTEAHVLLSGSQDGFMKCFDLRRKDSVSTFSGQSESVRDVQFSIRDYFTFA
SEG    ............................................................
1gotB  CCEEEEEEETT-TCEEEEEETTTEEEEEETTTTEEEEEECBTTCCEEEEEETTTTTEEEE

SEQ    STFENGNVQLWDIRRPDRCERMFTAHNGPVFCCDWHPEDRGWLATGGRDKMVKVWDMTTH
SEG    ............................................................
1gotB  E-ETTTEEEEEEETTTTEEEE-EEECCCCCEEEEEE-TTTTCCEEEEETTTEEEEEC....

SEQ    RAKEMHCVQTIASVARVKWRPECRHHLATCSMMVDHNIYVWDVRRPFVPAAMFEEHRDVT
SEG    ............................................................
1gotB  ............................................................

SEQ    TGIAWRHPHDPSFLLSGSKDSSLCQHLFRDASQPVERANPEGLCYGLFGDLAFAAKESLV
SEG    ............................................................
1gotB  ............................................................

SEQ    AAESGRKPYTGDRRHPIFFKRKLDPAEPFAGLASSALSVFETEPGGGGMRWFVDTAERYA
SEG    ............................................................
1gotB  ............................................................

SEQ    LAGRPLAELCDHNAKVARELGRNQVAQTWTMLRIIYCSPGLVPTANLNHSVGKGGSCGLP
SEG    ............................................................
1gotB  ............................................................

SEQ    LMNSFNLKDMAPGLGSETRLDRSKGDARSDTVLLDSSATLITNEDNEETEGSDVPADYLL
SEG    ........................................................xxxx
1gotB  ............................................................

SEQ    GDVEGEEDELYLLDPEHAHPEDPECVLPQEAFPLRHEIVDTPPGPEHLQDKADSPHVSGS
SEG    xxxxxxxxxxxxxx..............................................
1gotB  ............................................................

SEQ    EADVASLAPVDSSFSLLSVSHALYDSRLPPDFFGVLVRDMLHFYAEQGDVQMAVSVLIVL
SEG    ............................................................
1gotB  ............................................................

SEQ    GERVRKDIDEQTQEHWYTSYIDLLQRFRLWNVSNEVVKLSTSRAVSCLNQASTTLHVNCS
SEG    ............................................................
1gotB  ............................................................

SEQ    HCKRPMSSRGWVCDRCHRCASMCAVCHHVVKGLFVWCQGCSHGGHLQHIMKWLEGSSHCP
SEG    ............................................................
1gotB  ............................................................

SEQ    AGCGHLCEYS
SEG    ..........
1gotB  ..........
```

Prosite for DKFZphtes3_4f5.3

PS00001	74->78	ASN_GLYCOSYLATION	PDOC00001
PS00001	468->472	ASN_GLYCOSYLATION	PDOC00001
PS00001	691->695	ASN_GLYCOSYLATION	PDOC00001
PS00001	718->722	ASN_GLYCOSYLATION	PDOC00001
PS00004	69->73	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	152->156	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	17->20	PKC_PHOSPHO_SITE	PDOC00005
PS00005	165->168	PKC_PHOSPHO_SITE	PDOC00005
PS00005	172->175	PKC_PHOSPHO_SITE	PDOC00005
PS00005	239->242	PKC_PHOSPHO_SITE	PDOC00005
PS00005	364->367	PKC_PHOSPHO_SITE	PDOC00005
PS00005	701->704	PKC_PHOSPHO_SITE	PDOC00005

PS00005	727->730	PKC_PHOSPHO_SITE	PDOC00005
PS00006	76->80	CK2_PHOSPHO_SITE	PDOC00006
PS00006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	398->402	CK2_PHOSPHO_SITE	PDOC00006
PS00006	498->502	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	522->526	CK2_PHOSPHO_SITE	PDOC00006
PS00006	598->602	CK2_PHOSPHO_SITE	PDOC00006
PS00006	600->604	CK2_PHOSPHO_SITE	PDOC00006
PS00006	679->683	CK2_PHOSPHO_SITE	PDOC00006
PS00007	337->346	TYR_PHOSPHO_SITE	PDOC00007
PS00008	13->19	MYRISTYL	PDOC00008
PS00008	97->103	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	161->167	MYRISTYL	PDOC00008
PS00008	317->323	MYRISTYL	PDOC00008
PS00008	342->348	MYRISTYL	PDOC00008
PS00008	391->397	MYRISTYL	PDOC00008
PS00008	460->466	MYRISTYL	PDOC00008
PS00008	474->480	MYRISTYL	PDOC00008
PS00008	759->765	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009
PS00009	364->368	AMIDATION	PDOC00009
PS00190	743->749	CYTOCHROME_C	PDOC00169
PS00678	90->105	WD_REPEATS	PDOC00574
PS00678	223->238	WD_REPEATS	PDOC00574
PS00678	269->284	WD_REPEATS	PDOC00574

Pfam for DKFZphtes3_4f5.3

```
HMM_NAME     WD domain, G-beta repeats

HMM              *MrGHnnWVWCVaFSPDGrWFIvSGSWDgTCRLWD*
                  ++ HN++V C+ ++P+ R  +++G++D+ +++WD
Query        203  FTAHNGPVFCCDWHPEDRGWLATGGRDKMVKVWD    236
```

DKFZphtes3_4h6
----

group: intracellular transport/trafficking

DKFZphtes3_4h6 encodes a novel 622 amino acid protein with strong similarity to the kinesin light chain.

Kinesin is a microtubule-based motor protein that pulls vesicles or organelles towards the plus end of microtubules. Structural changes in the protein that drive motility are coupled to ATP binding and hydrolysis. The novel protein is similar to kinesin light chain, which is part of the functional kinesin holoenzyme tetrameric protein. The light chain has been proposed to function in coupling of cargo to the heavy chain or in the modulation of the ATPase activity of the heavy chain. The novel protein contains two kinesin light chain repeats and one RGD cell-attachment site.

The novel kinesin protein can find application in modulating the function of kinesin and modulating intracellular transport via/on microtubules.

strong similarity to Kinesin light chain

complete cDNA, complete cds, start at 150, EST hits (few)

Sequenced by AGOWA

Locus: unknown

Insert length: 2992 bp
Poly A stretch at pos. 2914, polyadenylation signal at pos. 2893

```
   1 GGCGGGATGG AGGCGGCGGG ACCGGCTCGC GGGTGCGGGT CCGGGTGAAG
  51 CGGGAGGCAG CCAGAGTCGG AGCCGGGCCC GAGCACCAGG CGCAGGCCCG
 101 GCGCCCGCCT GCCCGCACCC TCGTCCTCAC AGACGCCACA GCCATGGCCA
 151 TGATGGTGTT TCCGCGGGAG GAGAAGCTGA GCCAGGATGA GATCGTGCTG
 201 GGCACCAAGG CTGTCATCCA GGGACTGGAG ACTCTGCGTG GGGAGCATCG
 251 TGCCCTGCTG GCTCCTCTGG TTGCACCTGA GGCCGGCGAA GCCGAGCCTG
 301 GCTCGCAGGA GCGCTGCATC CTCCTGCGTC GCTCCCTGGA AGCCATTGAG
 351 CTTGGGCTGG GGGAGGCCCA GGTGATCTTG GCATTGTCGA GCCACCTGGG
 401 GGCTGTAGAA TCAGAGAAGC AGAAGCTGCG GGCGCAGGTG CGGCGTCTGG
 451 TGCAGGAGAA CCAGTGGCTG CGTGAGGAGC TGGCGGGGAC ACAGCAGAAG
 501 CTGCAGCGCA GTGAGCAGGC CGTGGCCCAG CTCGAGGAGG AGAAGCAGCA
 551 CTTGCTGTTC ATGAGCCAGA TCCGCAAGTT GGATGAAGAC GCCTCCCCTA
 601 ACGAGGAGAA GGGGGACGTC CCCAAAGACA CACTGGATGA CCTCTTCCCC
 651 AATGAGGATG AGCAGAGCCC AGCCCCTAGC CCAGGAGGAG GGGATGTGTC
 701 TGGTCAGCAT GGGGGCTACG AGATCCCGGC CCGGCTCCGC ACCCTGCACA
 751 ACCTGGTGAT CCAATACGCC TCACAGGGCC GCTACGAGGT AGCTGTGCCA
 801 CTCTGCAAGC AGGCACTCGA AGACCTGGAG AAGACGTCAG GCCACGACCA
 851 CCCTGACGTT GCCACCATGC TGAACATCCT GGCACTGGTC TATCGGGATC
 901 AGAACAAGTA CAAGGAGGCT GCCCACCTGC TCAATGATGC TCTGGCCATC
 951 CGGGAGAAAA CACTGGGCAA GGACCACCCA GCCGTGGCTG CGACACTAAA
1001 CAACCTGGCA GTCCTGTATG GCAAGAGGGG CAAGTACAAG GAGGCTGAGC
1051 CATTGTGCAA GCGGGCACTG GAGATCCGGG AGAAGGTCCT GGGCAAGTTT
1101 CACCCAGATG TGGCCAAGCA GCTCAGCAAC CTGGCCCTGC TGTGCCAGAA
1151 CCAGGGCAAA GCTGAGGAGG TGGAATATTA CTATCGGCGG GCACTGGAGA
1201 TCTATGCTAC ACGCCTCGGG CCCGATGACC CCAATGTGGC CAAGACCAAG
1251 AACAACCTGG CTTCCTGCTA CCTGAAGCAG GGCAAGTACC AGGATGCGGA
1301 GACCTTGTAC AAGGAGATCC TCACCCGCGC TCATGAGAAA GAGTTTGGCT
1351 CTGTCAATGG GGACAACAAG CCCATCTGGA TGCACGCAGA GGAGCGGGAG
1401 GAAAGCAAGG ATAAGCGCCG GGACAGCGCC CCCTATGGGG AATACGGCAG
1451 CTGGTACAAG GCCTGTAAAG TAGACAGCCC CACAGTCAAC ACCACCCTGC
1501 GCAGCTTGGG GGCCCTATAC CGGCGCCAGG GCAAGCTGGA AGCCGCGCAC
1551 ACACTAGAGG ACTGTGCCAG CCGTAACCGC AAGCAGGGTT TGGACCCCGC
1601 AAGCCAGACC AAGGTGGTAG AACTGCTGAA AGATGGCAGT GGCAGGCGGG
1651 GAGACCGCCG CAGCAGCCGA GACATGGCTG GGGGTGCCGG GCCTCGGTCT
1701 GAGTCTGACC TCGAGGACGT GGGACCTACA GCTGAGTGGA ATGGGGATGG
1751 CAGTGGCTCC TTGAGGCGCA GCGGTTCCTT TGGGAAACTC CGGGATGCCC
1801 TGAGGCGCAG CAGTGAGATG CTGGTAAAGA AGCTGCAGGG GGGCACCCCC
1851 CAGGAGCCCC CTAACCCCAG GATGAAGCGG GCCAGTTCCC TCAACTTCCT
1901 CAACAAGAGC GTGGAAGAGC CGACCCAGCC TGGAGGCACA GGTCTCTCTG
1951 ACAGCCGCAC TCTCAGCTCC AGCTCCATGG ACCTCTCCCG ACGAAGCTCC
2001 CTGGTGGGCT AATGCTGAAG GGGCAGCCAG TCACCAGAGC GCCCACCTGG
2051 CACACCCCCC TCACCCCAGC CCTGCGCATG GGCCTGCTGC TTGTCCCGCC
2101 TGTCTCTCCC ACAGCCCCTG TCTTTTCTGT TCAATCTCAG GGTAACCTTC
2151 TCCCTTGTCA TCTCAGCCTG AGCCCTGGAG GCTGGGCCTG CCCACTCCAG
2201 CTCCATCCCT TATTTATTCC TTCCAGCAGG GCCCTCTTCC CTAGGTTCGG
2251 GCCAGCAGGA GGTGCCGGCT GGAGTCTCCA CCATAGACTC AGTGGCCTGG
2301 CCTCCCCAGA CCCCAGAGCC AAGAACACTA AGCACTCGCC GGCCCTTCGG
2351 CACCCTCGCC CTCCCTCCCG ACTCAACCCG GCCGTTGCTT CTGTATATAG
2401 AGAAATAAGT TATTGGCCGC GCGCCTCCCT TCAGTCCACG GTACTACCCG
```

```
2451 GGCCTCCCCT CGTCCCTCTT CTAGTGGTAC CGCCCAGGCC TTAATCACCC
2501 CCATTCCGTG CGGTGGTATC TCCCAGGCTC TACATTCTCG GGAGCGGCGC
2551 CTCCCAAGGG GGTCCTGGGA CCTTCTCGCG CTCCTCCTGG CCTCTGAGGG
2601 ATGCGTCCTA CCCGCGCCAT CGCCCCGTGG CCCAGGACGG GGACCTCCCC
2651 TTAGTCCGTC CTCCCACCGC CGGGCCCTGC CCCGCATCCC GGCCTTATGC
2701 ACTGCCCCTC CCACCCGGCC CCGCCCAGGC ACGGCCGACC CCGCCCCGGG
2751 CACCGCCCAC CGAGCCATCC TGCCTCGCCT CCCCCCACGC CTGCAGCTTC
2801 TCGCGAGGGG CGGCGACGGT CCCCTGGTGG CAGGAGGGGC TCCCCCTGTT
2851 GCGGGTGAGG CGGCTGCTCT CTATTTTCAG ATGTTGCTGT AGAAATAAAG
2901 ACGGTTTAAA TCTGAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2951 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

98288268:
Two kinesin light chain genes in mice. Identification and characterization of the encoded proteins.

Peptide information for frame 3

ORF from 144 bp to 2009 bp; peptide length: 622
Category: strong similarity to known protein
Prosite motifs: RGD (502-505)
KINESIN_LIGHT (223-265)
KINESIN_LIGHT (265-307)

```
  1 MAMMVFPREE KLSQDEIVLG TKAVIQGLET LRGEHRALLA PLVAPEAGEA
 51 EPGSQERCIL LRRSLEAIEL GLGEAQVILA LSSHLGAVES EKQKLRAQVR
101 RLVQENQWLR EELAGTQQKL QRSEQAVAQL EEEKQHLLFM SQIRKLDEDA
151 SPNEEKGDVP KDTLDDLFPN EDEQSPAPSP GGGDVSGQHG GYEIPARLRT
201 LHNLVIQYAS QGRYEVAVPL CKQALEDLEK TSGHDHPDVA TMLNILALVY
251 RDQNKYKEAA HLLNDALAIR EKTLGKDHPA VAATLNNLAV LYGKRGKYKE
301 AEPLCKRALE IREKVLGKFH PDVAKQLSNL ALLCQNQGKA EEVEYYYRRA
351 LEIYATRLGP DDPNVAKTKN NLASCYLKQG KYQDAETLYK EILTRAHEKE
401 FGSVNGDNKP IWMHAEEREE SKDKRRDSAP YGEYGSWYKA CKVDSPTVNT
451 TLRSLGALYR RQGKLEAAHT LEDCASRNRK QGLDPASQTK VVELLKDGSG
501 RRGDRRSSRD MAGGAGPRSE SDLEDVGPTA EWNGDGSGSL RRSGSFGKLR
551 DALRRSSEML VKKLQGGTPQ EPPNPRMKRA SSLNFLNKSV EEPTQPGGTG
601 LSDSRTLSSS SMDLSRRSSL VG
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4h6, frame 3

TREMBL:AF055666_1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus kinesin light chain 2 (Klc2) mRNA, complete cds., N = 1, Score = 2824, P = 4e-294

PIR:I53013 kinesin light chain - human, N = 1, Score = 1927, P = 4.5e-199

PIR:C41539 kinesin light chain C - rat, N = 1, Score = 1919, P = 3.2e-198

SWISSPROT:KNLC_RAT KINESIN LIGHT CHAIN (KLC)., N = 1, Score = 1919, P = 3.2e-198

>TREMBL:AF055666_1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus kinesin light chain 2 (Klc2) mRNA, complete cds.
Length = 599

HSPs:

```
Score = 2824 (423.7 bits), Expect = 4.0e-294, P = 4.0e-294
Identities = 558/598 (93%), Positives = 572/598 (95%)

Query:    1 MAMMVFPREEKLSQDEIVLGTKAVIQGLETLRGEHRALLAPLVAPEAGEAEPGSQERCIL 60
            MA MV PREEKLSQDEIVLGTKAVIQGLETLRGEHRALLAPL + EAGEAEPGSQERC+L
Sbjct:    1 MATMVLPREEKLSQDEIVLGTKAVIQGLETLRGEHRALLAPLASHEAGEAEPGSQERCLL 60

Query:   61 LRRSLEAIELGLGEAQVILALSSHLGAVESEKQKLRAQVRRLVQENQWLREELAGTQQKL 120
            LRRSLEAIELGLGEAQVILALSSHLGAVESEKQKLRAQVRRLVQENQWLREELAGTQQKL
Sbjct:   61 LRRSLEAIELGLGEAQVILALSSHLGAVESEKQKLRAQVRRLVQENQWLREELAGTQQKL 120

Query:  121 QRSEQAVAQLEEEKQHLLFMSQIRKLDEDASPNEEKGDVPKDTLDDLFPNEDEQSPAPSP 180
            QRSEQAVAQLEEEKQHLLFMSQIRKLDE    P EEKGDVPKD+LDDLFPNEDEQSPAPSP
Sbjct:  121 QRSEQAVAQLEEEKQHLLFMSQIRKLDE-MLPQEEKGDVPKDSLDDLFPNEDEQSPAPSP 179

Query:  181 GGGDVSGQHGGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALEDLEKTSGHDHPDVA 240
            GGGDV+ QHGGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALEDLEKTSGHDHPDVA
Sbjct:  180 GGGDVAAQHGGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALEDLEKTSGHDHPDVA 239

Query:  241 TMLNILALVYRDQNKYKEAAHLLNDALAIREKTLGKDHPAVAATLNNLAVLYGKRGKYKE 300
            TMLNILALVYRDQNKYK+AAHLLNDALAIREKTLGKDHPAVAATLNNLAVLYGKRGKYKE
Sbjct:  240 TMLNILALVYRDQNKYKDAAHLLNDALAIREKTLGKDHPAVAATLNNLAVLYGKRGKYKE 299

Query:  301 AEPLCKRALEIREKVLGKFHPDVAKQLSNLALLCQNQGKAEEVEYYYRRALEIYATRLGP 360
            AEPLCKRALEIREKVLGKFHPDVAKQLSNLALLCQNQGKAEEVEYYYRRALEIYATRLGP
Sbjct:  300 AEPLCKRALEIREKVLGKFHPDVAKQLSNLALLCQNQGKAEEVEYYYRRALEIYATRLGP 359

Query:  361 DDPNVAKTKNNLASCYLKQGKYQDAETLYKEILTRAHEKEFGSVNGDNKPIWMHAEEREE 420
            DDPNVAKTKNNLASCYLKQGKYQDAETLYKEILTRAHEKEFGSVNG+NKPIWMHAEEREE
Sbjct:  360 DDPNVAKTKNNLASCYLKQGKYQDAETLYKEILTRAHEKEFGSVNGENKPIWMHAEEREE 419

Query:  421 SKDKRRDSAPYGEYGSWYKACKVDSPTVNTTLRSLGALYRRQGKLEAAHTLEDCASRNRK 480
            SKDKRRD  P   EYGSWYKACKVDSPTVNTTLR+LGALYR +GKLEAAHTLEDCASR+RK
Sbjct:  420 SKDKRRDRRPM-EYGSWYKACKVDSPTVNTTLRTLGALYRPEGKLEAAHTLEDCASRSRK 478

Query:  481 QGLDPASQTKVVELLKDGSGRRGDRRSSRDMAGGAGPRSESDLEDVGPTAEWNGDGSGSL 540
            QGLDPASQTKVVELLKDGSGR G RR SRD+AG   P+SESDLE+ GP AEW+GDGSGSL
Sbjct:  479 QGLDPASQTKVVELLKDGSGR-GHRRGSRDVAG---PQSESDLEESGPAAEWSGDGSGSL 534

Query:  541 RRSGSFGKLRDALRRSSEMLVKKLQGGTPQEPPNPRMKRASSLNFLNKSVEEPTQPGG 598
            RRSGSFGKLRDALRRSSEMLV+KLQGG PQEP N RMKRASSLNFLNKSVEEP QPGG
Sbjct:  535 RRSGSFGKLRDALRRSSEMLVRKLQGGGPQEP-NSRMKRASSLNFLNKSVEEPVQPGG 591
```

Pedant information for DKFZphtes3_4h6, frame 3

Report for DKFZphtes3_4h6.3

```
[LENGTH]     622
[MW]         68934.82
[pI]         6.72
[HOMOL]      TREMBL:AF055666_1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus
kinesin light chain 2 (Klc2) mRNA, complete cds. 0.0
[BLOCKS]     BL00927C Trehalase proteins
[BLOCKS]     BL01160I Kinesin light chain repeat proteins
[BLOCKS]     BL01160H Kinesin light chain repeat proteins
[BLOCKS]     BL01160G Kinesin light chain repeat proteins
[BLOCKS]     BL01160F Kinesin light chain repeat proteins
[BLOCKS]     BL01160E Kinesin light chain repeat proteins
[BLOCKS]     BL01160D Kinesin light chain repeat proteins
[BLOCKS]     BL01160C Kinesin light chain repeat proteins
[BLOCKS]     BL01160B Kinesin light chain repeat proteins
[BLOCKS]     BL01160A Kinesin light chain repeat proteins
[SUPFAM]     tetratricopeptide repeat homology 1e-07
[PROSITE]    RGD          1
[PROSITE]    MYRISTYL     8
[PROSITE]    KINESIN_LIGHT  2
[PROSITE]    AMIDATION    2
[PROSITE]    CAMP_PHOSPHO_SITE   5
[PROSITE]    CK2_PHOSPHO_SITE    11
[PROSITE]    TYR_PHOSPHO_SITE    3
[PROSITE]    PKC_PHOSPHO_SITE    7
[PROSITE]    ASN_GLYCOSYLATION   2
[PFAM]       Kinesin light chain repeat
[KW]         All_Alpha
[KW]         LOW_COMPLEXITY    12.54 %
[KW]         COILED_COIL        4.98 %
```

```
SEQ     MAMMVFPREEKLSQDEIVLGTKAVIQGLETLRGEHRALLAPLVAPEAGEAEPGSQERCIL
SEG     ............................................................
PRD     cccchhhhhhhhhhhhchhhhhhhhhhhhhhhchhhhhhhhhhhhhhcccccccchhhhh
COILS   ............................................................

SEQ     LRRSLEAIELGLGEAQVILALSSHLGAVESEKQKLRAQVRRLVQENQWLREELAGTQQKL
SEG     ............................................................
PRD     hhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhh
COILS   .................................................CCCCCCCCCCC

SEQ     QRSEQAVAQLEEEKQHLLFMSQIRKLDEDASPNEEKGDVPKDTLDDLFPNEDEQSPAPSP
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccc
COILS   CCCCCCCCCCCCCCCCC...........................................

SEQ     GGGDVSGQHGGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALEDLEKTSGHDHPDVA
SEG     ............................................................
PRD     cccccccccccccchhhhhhhhhhhhhhccceeeeeehhhhhhhhhhhhhhccccccchh
COILS   ............................................................

SEQ     TMLNILALVYRDQNKYKEAAHLLNDALAIREKTLGKDHPAVAATLNNLAVLYGKRGKYKE
SEG     .......................................xxxxxxxxxxxx.........
PRD     hhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhccccccchhhhhhhhhhhhhhcccccchh
COILS   ............................................................

SEQ     AEPLCKRALEIREKVLGKFHPDVAKQLSNLALLCQNQGKAEEVEYYYRRALEIYATRLGP
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhccc
COILS   ............................................................

SEQ     DDPNVAKTKNNLASCYLKQGKYQDAETLYKEILTRAHEKEFGSVNGDNKPIWMHAEEREE
SEG     .......................................................xxxxx
PRD     ccccccchhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhcccccccccchhhhhhhhhh
COILS   ............................................................

SEQ     SKDKRRDSAPYGEYGSWYKACKVDSPTVNTTLRSLGALYRRQGKLEAAHTLEDCASRNRK
SEG     xxxxxxxx....................................................
PRD     hhhhcccccccccccccceeeecccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   ............................................................

SEQ     QGLDPASQTKVVELLKDGSGRRGDRRSSRDMAGGAGPRSESDLEDVGPTAEWNGDGSGSL
SEG     ................xxxxxxxxxxxxxx.........................xxxxx
PRD     hhccchhhhhhhhhhhccccccccccccccccccccccccccccceeeeccccccccccc
COILS   ............................................................

SEQ     RRSGSFGKLRDALRRSSEMLVKKLQGGTPQEPPNPRMKRASSLNFLNKSVEEPTQPGGTG
SEG     xxxxxxxxxx..............................................xxxx
PRD     ccccccchhhhhhhhhhhhhhhhhcccccccccccchhhhhhccccccccccccccccccc
COILS   ............................................................

SEQ     LSDSRTLSSSSMDLSRRSSLVG
SEG     xxxxxxxxxxxxxxxxxxxx..
PRD     cccccccccccchhhhhhcccc
COILS   ......................
```

Prosite for DKFZphtes3_4h6.3

PS00001	449->453	ASN_GLYCOSYLATION	PDOC00001
PS00001	587->591	ASN_GLYCOSYLATION	PDOC00001
PS00004	425->429	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	505->509	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	554->558	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	578->582	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	616->620	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	90->93	PKC_PHOSPHO_SITE	PDOC00005
PS00005	451->454	PKC_PHOSPHO_SITE	PDOC00005
PS00005	499->502	PKC_PHOSPHO_SITE	PDOC00005
PS00005	507->510	PKC_PHOSPHO_SITE	PDOC00005
PS00005	539->542	PKC_PHOSPHO_SITE	PDOC00005
PS00005	615->618	PKC_PHOSPHO_SITE	PDOC00005
PS00006	13->17	CK2_PHOSPHO_SITE	PDOC00006
PS00006	151->155	CK2_PHOSPHO_SITE	PDOC00006
PS00006	163->167	CK2_PHOSPHO_SITE	PDOC00006
PS00006	232->236	CK2_PHOSPHO_SITE	PDOC00006
PS00006	470->474	CK2_PHOSPHO_SITE	PDOC00006
PS00006	507->511	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	521->525	CK2_PHOSPHO_SITE	PDOC00006

PS00006	568->572	CK2_PHOSPHO_SITE	PDOC00006
PS00006	589->593	CK2_PHOSPHO_SITE	PDOC00006
PS00006	610->614	CK2_PHOSPHO_SITE	PDOC00006
PS00007	339->346	TYR_PHOSPHO_SITE	PDOC00007
PS00007	339->347	TYR_PHOSPHO_SITE	PDOC00007
PS00007	424->432	TYR_PHOSPHO_SITE	PDOC00007
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	86->92	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	187->193	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	482->488	MYRISTYL	PDOC00008
PS00008	598->604	MYRISTYL	PDOC00008
PS00008	600->606	MYRISTYL	PDOC00008
PS00009	292->296	AMIDATION	PDOC00009
PS00009	499->503	AMIDATION	PDOC00009
PS00016	502->505	RGD	PDOC00016
PS01160	223->265	KINESIN_LIGHT	PDOC00893
PS01160	265->307	KINESIN_LIGHT	PDOC00893

Pfam for DKFZphtes3_4h6.3

```
HMM_NAME       Kinesin light chain repeat

HMM                 *RALEDREKtlGHDHPDVAtMLNNLALvCRNQNKYeEveNYYN*
                     +ALED+EKT+GHDHPDVATMLN+LALV+R+QNKY+E++ ++N
Query         223   QALEDLEKTSGHDHPDVATMLNILALVYRDQNKYKEAAHLLN    264

50.46     265   306     1    42 dkfzphtes3_4h6.3 strong similarity to Kinesin light chain
  Alignment to HMM consensus:
Query               *RALEDREKtlGHDHPDVAtMLNNLALvCRNQNKYeEveNYYN*
                      AL +REKTLG DHP VA  LNNLA+++  ++KY+E+E + +
  dkfzphtes3  265   DALAIREKTLGKDHPAVAATLNNLAVLYGKRGKYKEAEPLCK    306

Query               348     1    42 dkfzphtes3_4h6.3 strong similarity to Kinesin light chain
  Alignment to HMM consensus:
HMM                 *RALEDREKtlGHDHPDVAtMLNNLALvCRNQNKYeEveNYYN*
                     RALE+REK+LG  HPDVA++L+NLAL+C+NQ+K EEVE YY+
Query         307   RALEIREKVLGKFHPDVAKQLSNLALLCQNQGKAEEVEYYYR    348

39.10     349   390     1    42 dkfzphtes3_4h6.3 strong similarity to Kinesin light chain
  Alignment to HMM consensus:
Query               *RALEDREKtlGHDHPDVAtMLNNLALvCRNQNKYeEveNYYN*
                     RALE+     LG D P+VA+  NNLA  +  Q+KY+++E +Y+
  dkfzphtes3  349   RALEIYATRLGPDDPNVAKTKNNLASCYLKQGKYQDAETLYK    390
```

DKFZphtes3_4o19
----

group: testes derived

DKFZphtes3_4o19 encodes a novel 1180 amino acid protein with weak similarity to human megakaryocyte stimulating factor and human mucin.

The novel protein contains a cytochrome c family heme-binding site signature.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to megakaryocyte stimulating factor and mucin

complete cDNA, complete cds, EST hits (few)

Sequenced by AGOWA

Locus: unknown

Insert length: 3767 bp
Poly A stretch at pos. 3757, polyadenylation signal at pos. 3737

```
   1 GGCTAGGTTT AGCTTCAGGG GCAGCCCAGG GCAGTGTTGC TGCATATTGC
  51 ATGGATGAAA GGCTGAAGGC TGCCTCCTCT TGCAGGCTGG CTTCTGAGAT
 101 TGCACCTTCT TCTCCTGCTA CTCCTCCAAA TCTATGACCC TTCAAGGCAG
 151 AGCTGACCTG TCCGGTAATC AAGGCAATGC AGCCGGCCGC CTAGCTACAG
 201 TTCACGAGCC AGTTGTCACC CAGTGGGCGG TGCATCCTCC AGCCCCCGCT
 251 CACCCCAGTC TCCTGGACAA AATGGAGAAA GCGCCTCCAC AGCCCCAGCA
 301 CGAGGGCCTC AAGTCCAAGG AGCATCTTCC GCAACAGCCT GCCGAAGGCA
 351 AGACGGCGTC CCGCCGCGTC CCACGCCTCC GGGCTGTGGT CGAGAGCCAG
 401 GCCTTCAAGA ACATCCTGGT AGACGAGATG GACATGATGC ACGCCCGTGC
 451 AGCCACGCTC ATCCAAGCCA ACTGGAGGGG CTATTGGCTC CGGCAGAAGC
 501 TGATTTCCCA GATGATGGCG GCCAAGGCCA TCCAGGAGGC CTGGCGGCGC
 551 TTCAACAAGA GACACATCCT TCACTCCAGC AAGTCGTTGG TAAAGAAAAC
 601 GAGGGCGGAG GAGGGGGACA TACCTTATCA CGCCCCACAG CAGGTGCGCT
 651 TCCAGCATCC GGAAGAGAAC CGCCTTCTGT CCCCGCCCAT CATGGTGAAC
 701 AAGGAGACCC AGTTCCCTTC CTGTGACAAT CTGGTCCTCT GCAGACCCCA
 751 GTCGTCCCCC CTCCTGCAGC CCCCAGCAGC TCAGGGTACC CCAGAGCCCT
 801 GTGTGCAGGG TCCTCATGCT GCCAGAGTCC GGGGGCTGGC CTTCCTGCCA
 851 CACCAGACGG TCACCATCAG ATTTCCCTGC CCAGTGAGTT TGGACGCAAA
 901 ATGCCAGCCA TGCCTGCTGA CCAGAACCAT CAGAAGCACC TGCCTCGTCC
 951 ACATAGAGGG TGACTCAGTG AAGACCAAAC GTGTAAGTGC CCGGACCAAC
1001 AAAGCCAGGG CTCCGGAGAC ACCATTGTCC AGAAGGTATG ACCAGGCAGT
1051 TACGAGACCA TCCAGAGCCC AAACCCAGGG CCCTGTGAAA GCAGAGACCC
1101 CCAAAGCCCC CTTCCAGATA TGTCCAGGGC CCATGATCAC CAAGACTCTA
1151 CTCCAGACAT ATCCAGTGGT CTCCGTGACC CTGCCACAGA CATATCCAGC
1201 GTCCACGATG ACCACCACCC CACCCAAGAC TAGCCCAGTT CCCAAAGTAA
1251 CAATAATCAA GACCCCAGCC CAGATGTATC CGGGGCCCAC AGTGACCAAA
1301 ACTGCACCTC ACACATGCCC CATGCCCACA ATGACCAAGA TCCAGGTACA
1351 CCCCACAGCC TCCAGAACTG GCACCCCACG GCAGACATGC CCTGCGACCA
1401 TCACGGCAAA GAACCGACCT CAGGTTTCCC TTCTGGCTTC CATCATGAAG
1451 AGCCTGCCCC AGGTATGCCC GGGGCCTGCG ATGGCAAAGA CCCCACCCCA
1501 GATGCACCCG GTCACCACCC CAGCCAAAAA CCCATTGCAA ACATGTCTGT
1551 CAGCCACAAT GTCCAAGACT TCATCCCAGA GGAGCCCAGT TGGGGTGACC
1601 AAGCCCTCAC CCCAGACCCG CCTGCCAGCC ATGATAACCA AGACCCCAGC
1651 CCAGTTACGC TCGGTGGCCA CCATCCTCAA GACTCTGTGT CTGGCCTCTC
1701 CAACAGTGGC AAATGTCAAG GCTCCACCCC AAGTGGCGGT AGCAGCCGGA
1751 ACTCCCAACA CCTCAGGCTC CATCCATGAG AACCCACCCA AGGCCAAGGC
1801 CACCGTGAAT GTGAAGCAGG CTGCAAAGGT GGTGAAAGCC TCATCCCCCT
1851 CCTATTTGGC TGAGGGGAAG ATCAGGTGCC TGGCTCAACC ACATCCGGGA
1901 ACTGGGGTCC CCAGGGCTGC AGCTGAGCTT CCTTTGGAAG CCGAGAAAAT
1951 CAAGACTGGC ACCCAGAAAC AGGCGAAAAC AGACATGGCA TTTAAGACCA
2001 GTGTGGCAGT GGAAATGGCT GGGGCTCCAT CCTGGACAAA AGTTGCTGAG
2051 GAAGGGGACA AGCCACCTCA CGTGTATGTG CCTGTAGACA TGGCTGTCAC
2101 CCTGCCCCGG GGACAGCTGG CTGCCCCACT GACCAATGCC TCATCCCAGA
2151 GACATCCACC CTGCCTGTCC CAGAGACCAC TGGCCGCCCC GCTGACCAAG
2201 GCCTCATCTC AGGGACATCT GCCCACTGAG CTGACCAAGA CCCCATCCCT
2251 GGCCCATCTG GACACCTGTC TGAGCAAGAT GCATTCCCAG ACACATCTGG
2301 CCACAGGTGC CGTGAAGGTC CAGTCCCAAG CGCCTCTAGC CACCTGTCTG
2351 ACCAAGACGC AGTCCCGGGG GCAGCCGATC ACAGACATAA CCACGTGCCT
2401 CATCCCAGCG CACCAGGCTG CTGATCTCAG CAGCAACACC CACTCCCAGG
2451 TGCTCCTAAC AGGGTCCAAG GTGTCCAACC ACGCCTGCCA GCGCCTCGGT
2501 GGCCTCAGCG CCCCACCCTG GGCCAAGCCA GAGGACAGAC AGACCCAGCC
2551 ACAGCCCCAC GGACACGTGC CGGGGAAGAC CACTCAGGGG GGACCATGCC
2601 CGGCAGCCTG TGAGGTCCAG GGTATGCTGG TGCCGCCGAT GGCACCCACC
```

```
2651 GGCCATTCCA CATGCAACGT TGAGTCCTGG GGAGACAACG GAGCCACACG
2701 TGCCCAGCCA TCAATGCCCG GCCAGGCGGT GCCCTGCCAG GAGGACACGG
2751 GCCCCGCGGA CGCTGGTGTG GTTGGTGGCC AATCGTGGAA CCGCGCATGG
2801 GAGCCAGCCA GGGGTGCTGC GTCCTGGGAC ACCTGGCGCA ACAAGGCGGT
2851 GGTGCCTCCC AGGCGGTCCG GGGAGCCAAT GGTGTCCATG CAGGCTGCAG
2901 AGGAGATCCG CATCCTCGCA GTGATCACTA TCCAGGCGGG CGTCCGTGGC
2951 TACCTGGCGC GTCGCAGGAT CCGGCTGTGG CACCGGGGGG CCATGGTCAT
3001 CCAAGCTACT TGGCGCGGCT ACCGTGTGCG GCGGAACCTG GCACACCTCT
3051 GCAGAGCCAC CACGACCATC CAGTCTGCCT GGCGCGGCTA CAGCACCCGC
3101 CGGGACCAAG CCCGGCACTG GCAGATGCTC CACCCCGTCA CGTGGGTGGA
3151 GCTGGGCAGC CGGGCCGGGG TCATGTCTGA CCGAAGCTGG TTCCAGGATG
3201 GCAGAGCCAG GACAGTATCT GACCATCGCT GCTTCCAGTC CTGCCAGGCA
3251 CACGCTTGCA GCGTCTGCCA CTCCCTGAGC TCCAGGATCG GGAGCCCGCC
3301 CAGCGTGGTG ATGCTAGTGG GCTCCAGCCC TCGCACCTGT CATACCTGTG
3351 GACGCACACA GCCCACCCGT GTGGTGCAGG GCATGGGCCA GGGCACTGAG
3401 GGCCCCGGGG CAGTGTCTTG GGCCTCCGCC TACCAGCTGG CTGCCCTGAG
3451 TCCCAGGCAG CCGCATCGCC AGGACAAAGC GGCCACAGCC ATCCAGTCCG
3501 CCTGGAGGGG CTTTAAGATC CGCCAGCAGA TGAGGCAGCA GCAAATGGCA
3551 GCGAAGATAG TTCAAGCCAC CTGGCGAGGC CACCATACCC GGAGCTGTCT
3601 GAAGAACACA GAGGCCTCT TGGGACCAGC AGACCCCTCG GCCAGCTCAC
3651 GGCACATGCA TTGGCCTGGC ATCTAGGACC CTGGCTCCCT GCAGTGGGGA
3701 CTTCGTGGGA GGCACTCATG GCTCTCTGGG TCTAATGAAT AAAGTCCTCC
3751 ACAGCCTAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 3673 bp; peptide length: 1180
Category: similarity to known protein

```
   1 MTLQGRADLS GNQGNAAGRL ATVHEPVVTQ WAVHPPAPAH PSLLDKMEKA
  51 PPQPQHEGLK SKEHLPQQPA EGKTASRRVP RLRAVVESQA FKNILVDEMD
 101 MMHARAATLI QANWRGYWLR QKLISQMMAA KAIQEAWRRF NKRHILHSSK
 151 SLVKKTRAEE GDIPYHAPQQ VRFQHPEENR LLSPPIMVNK ETQFPSCDNL
 201 VLCRPQSSPL LQPPAAQGTP EPCVQGPHAA RVRGLAFLPH QTVTIRFPCP
 251 VSLDAKCQPC LLTRTIRSTC LVHIEGDSVK TKRVSARTNK ARAPETPLSR
 301 RYDQAVTRPS RAQTQGPVKA ETPKAPFQIC PGPMITKTLL QTYPVVSVTL
 351 PQTYPASTMT TTPPKTSPVP KVTIIKTPAQ MYPGPTVTKT APHTCPMPTM
 401 TKIQVHPTAS RTGTPRQTCP ATITAKNRPQ VSLLASIMKS LPQVCPGPAM
 451 AKTPPQMHPV TTPAKNPLQT CLSATMSKTS SQRSPVGVTK PSPQTRLPAM
 501 ITKTPAQLRS VATILKTLCL ASPTVANVKA PPQVAVAAGT PNTSGSIHEN
 551 PPKAKATVNV KQAAKVVKAS SPSYLAEGKI RCLAQPHPGT GVPRAAAELP
 601 LEAEKIKTGT QKQAKTDMAF KTSVAVEMAG APSWTKVAEE GDKPPHVYVP
 651 VDMAVTLPRG QLAAPLTNAS SQRHPPCLSQ RPLAAPLTKA SSQGHLPTEL
 701 TKTPSLAHLD TCLSKMHSQT HLATGAVKVQ SQAPLATCLT KTQSRGQPIT
 751 DITTCLIPAH QAADLSSNTH SQVLLTGSKV SNHACQRLGG LSAPPWAKPE
 801 DRQTQPQPHG HVPGKTTQGG PCPAACEVQG MLVPPMAPTG HSTCNVESWG
 851 DNGATRAQPS MPGQAVPCQE DTGPADAGVV GGQSWNRAWE PARGAASWDT
 901 WRNKAVVPPR RSGEPMVSMQ AAEEIRILAV ITIQAGVRGY LARRRIRLWH
 951 RGAMVIQATW RGYRVRRNLA HLCRATTTIQ SAWRGYSTRR DQARHWQMLH
1001 PVTWVELGSR AGVMSDRSWF QDGRARTVSD HRCFQSCQAH ACSVCHSLSS
1051 RIGSPPSVVM LVGSSPRTCH TCGRTQPTRV VQGMGQGTEG PGAVSWASAY
1101 QLAALSPRQP HRQDKAATAI QSAWRGFKIR QQMRQQQMAA KIVQATWRGH
1151 HTRSCLKNTE ALLGPADPSA SSRHMHWPGI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_4o19, frame 2

TREMBL:HSU70136_1 product: "megakaryocyte stimulating factor"; Human megakaryocyte stimulating factor mRNA, complete cds., N = 2, Score = 242, P = 9.6e-16

TREMBL:HSMUC2A_1 gene: "MUC2"; product: "mucin"; Human mucin-2 gene, partial cds., N = 1, Score = 204, P = 1.4e-12

PIR:S48478 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae), N = 1, Score = 192, P = 9.6e-11

>TREMBL:HSU70136_1 product: "megakaryocyte stimulating factor"; Human megakaryocyte stimulating factor mRNA, complete cds.
Length = 1,404

HSPs:

Score = 242 (36.3 bits), Expect = 9.6e-16, Sum P(2) = 9.6e-16
Identities = 145/546 (26%), Positives = 198/546 (36%)

```
Query:   282 KRVSARTNKARAPETPLSRRYDQAVTRPSRAQTQGPVKAETPKAPFQIC-PGPMITKTLL 340
             K+ +  T K  AP TP             PS    + P     T  AP     P P  TK+
Sbjct:   488 KKPAPTTPKEPAPTTP-KEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTKSAP 546

Query:   341 QTYPVVSVTLPQ----TYPASTMTTTPPKTSPV-PKVTIIKTPAQMYPGPTVTKTAPHTC 395
             T    S T  +     T P    TTP K +P  PK     TP +  P PT TK
Sbjct:   547 TTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKE--PAPTTTKK----- 599

Query:   396 PMPTMTKIQVHPTASRTGTPRQTCPATITAKNRPQVSLLASIMKSLPQVCPGPAMAKTPP 455
             P PT  K +  PT      TP++T P T                LA      P       +A T P
Sbjct:   600 PAPTAPK-EPAPT-----TPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTP 653

Query:   456 QMHPVTTPAKNPLQTCLSATMSKTSSQRSPVGVTKPSPQT-RLPAMIT-KTPAQLRSVAT 513
             +      TTP + P  T    A      T  + +P     +P+P T + PA  T K  A       T
Sbjct:   654 EEPTPTTP-EEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGT 712

Query:   514 ILKTLCLASPTVANVKAPPQVAVAAG---TPNTSGSIHENPPKAKATVNVKQAAKVV-KA 569
              TL    +PT       AP ++A        T  TS            PK  A    K+ A    K
Sbjct:   713 APTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKE 772

Query:   570 SSPSYLAEGKIRCLAQPHPGTGVPRAAAELPLEAEKIKTGT--QKQAKTDMAFKTSVAVE 627
             +P+           L +P P T     A  EL           K  T T       K A T    +T+
Sbjct:   773 PAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKGPTSTTSDKPAPTTPK-ETAPTTP 831

Query:   628 MAGAPSWTKVAEEGDKPPHVYVPVDMAVTLPRGQLAAPLTNASSQRHPPCLSQRPLAAPL 687
                AP+  K   +      P     P     V+ P         +  S         P LS  P     L
Sbjct:   832 KEPAPTTPK--KPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKAL 889

Query:   688 TKASSQGHLPTELTKTPSLA--HLDTCLSKMHSQTHLATGAVKVQSQAPLAT--CLTKTQ 743
              +  +  +PT  TKTP+      +  T        ++  L T    +  + AP  T     T T+
Sbjct:   890 ENSPKEPGVPT--TKTPAATKPEMTTTAKDKTTERDLRT-TPETTTAAPKMTKETATTTE 946

Query:   744 SRGQPITDITTCLIPAHQAADLS--SNTHSQVLLTGSKVSN--HACQRLGGLSAPP-WAK 798
                +           TT  + +     D +     T  +         KV+             ++ P    AK
Sbjct:   947 KTTESKITATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAK 1006

Query:   799 PEDRQTQPQPHGHVPGKTTQGGPCPAA 825
             P+DR T  +    P K T+   P +
Sbjct:  1007 PKDRATNSKATTPKPQKPTKAPKKPTS 1033
```

Score = 205 (30.8 bits), Expect = 3.1e-12, Sum P(2) = 3.1e-12
Identities = 146/565 (25%), Positives = 209/565 (36%)

```
Query:   281 TKRVSARTNKARAPETPLSRRYDQAVTRPSRAQTQGPVKAE--TPKAPFQICPGPMITKT 338
             TK+ +    K  AP TP       + A T P +     P K    TP+ P     P  +   T
Sbjct:   597 TKKPAPTAPKEPAPTTPK----ETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT 652

Query:   339 LLQTYPVVSVTLPQTYPASTMTTTPPKTSPV-PKVTIIKTPAQMYPGPTVTK-TAPHTCP 396
             + P       T P +    TP + +P  PK      TP +  P PT  K TAP T P
Sbjct:   653 PEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKE--PAPTTPKETAP-TTP 709

Query:   397 M---PTMTKIQVHPTASRTGTPRQTCPATITAKNRPQVSLLASIMKSLPQVCPGPAMAKT 453
                 PT  K +  PT    P++  P  T       +     S   +      K P      G A    T
Sbjct:   710 KGTAPTTLK-EPAPTTPKKPAPKELAPTT----TKEPTSTTSD--KPAPTTPKGTAPT-T 761

Query:   454 PPQMHPVTTPAKNPLQTCLSATMSKTSSQRSPVGVTKPSPQTRLPAMITKTPAQLRSVAT 513
             P +  P TTP K P  T      T   T  + +P    KP+P+  P     TK P     S
Sbjct:   762 PKEPAP-TTP-KEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTT-TKGPTSTTSDKP 818

Query:   514 ILKTLCLASPTVANVKAPPQVAVAAGTPNTSGSIHENPPKAKATVNV----KQAAKVVKA 569
                T    +PT       AP       A  P T        E PP    + V+       K+    + K+
Sbjct:   819 APTTPKETAPTTPKEPAPTTPKKPA--PTTP----ETPPPTTSEVSTPTTTKEPTTIHKS 872

Query:   570 ---SSPSYLAEGKIRCLAQPHPGTGVPRAAAELPLEAEKIKTGTQKQAKTDMAFKTSVAV 626
                S+P    AE    + L         GVP        + P      +   T  T K       T+     +T+
```

```
Sbjct:  873 PDESTPELSAEPTPKALENSPKEPGVP--TTKTPAATKPEMTTTAKDKTTERDLRTTPET 930

Query:  627 EMAGAPSWTK-VAEEGDKPPHVYVPVDMAVTLPRGQLAAPLTNASSQRHPPCLSQRPLAA 685
              A AP   TK  A   +K        +  +T    Q+ +  T  ++         L  LA
Sbjct:  931 TTA-APKMTKETATTTEKT------TESKITATTTQVTSTTTQDTTPFKITTLKTTTLAP 983

Query:  686 PLTKASSQGHLPTELTKTPSLAHLDTCLSKMHSQTHLATGAVKVQS-----QAPLATCLT 740
             +T  + +    TE+   P     +T   K  +    AT   K Q      + P +T
Sbjct:  984 KVT-TTKKTITTTEIMNKPE----ETAKPKDRATNSKAT-TPKPQKPTKAPKKPTSTKKP 1037

Query:  741 KTQSR-GQPITDIT----TCLIPAHQAADLSSNTHSQVLLTGSKVSNHACQRLGGLSAPP 795
            KT   R  +P T  T    T  +P            +    Q      ++  N     +  S
Sbjct: 1038 KTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDA 1097

Query:  796 W-AKPEDRQTQPQPHGHVPGKTTQGGPCPAACEVQGMLVPPMAPTGHSTCN 845
              A+ E        +PH  +P  T        P     QG+++ PM      + CN
Sbjct: 1098 GGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPN-QGIIINPMLSDETNICN 1147

 Score = 198 (29.7 bits), Expect = 2.3e-11, Sum P(2) = 2.3e-11
 Identities = 142/513 (27%), Positives = 200/513 (38%)

Query:  204 RPQSSPLLQPPAAQGTPEPCVQGPHAARVRGLAFLPHQTVTIRFPCPVSLDAKCQPCLLT 263
            R +   P  +PP                 G           +  H  V+          +    +P L
Sbjct:  207 RTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSLPP 266

Query:  264 R--TIRSTCLVHIEGDSVKTKRVSARTNKARAP---ETPLSRRYDQAVTRPSR---AQTQ 315
               T + T L    +  +V+TK  +  TNK  +     E   S +  Q++ + S    A T
Sbjct:  267 NSDTSKETSLTVNKETTVETKETTT-TNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTS 325

Query:  316 GPVKAETPKAPFQICPGPMITKTLLQTYPVVSVTLPQTYPASTMTTTPPKTSPVPKVTII 375
              +    TPKA        GP +T T  +  P      T P+  PAST   TP + +P     +
Sbjct:  326 KVLAKPTPKAE-TTTKGPALT-TPKEPTP----TTPKE-PAST---TPKEPTPTTIKSAP 375

Query:  376 KTPAQMYPGPTVTKTAPHTC--PMPTMTKIQVHPTASRTGTPRQTC-PATITAKNRPQVS 432
             TP +   P PT TK+AP T    P PT TK +  PT  +   P  T  PA  T K+ P
Sbjct:  376 TTPKE--PAPTTTKSAPTTPKEPAPTTTK-EPAPTTPKEPAPTTTKEPAPTTTKSAPTTP 432

Query:  433 ---LLASIMKSLPQVCPGPAMAKTPPQMHPVTTPAKNPLQTCLSATMSKTSSQRSPVGVT 489
                   + K  P    PA   TP +  P TTP K P  T        + T  + +P
Sbjct:  433 KEPAPTTPKKPAPTTPKEPAPT-TPKEPTP-TTP-KEPAPTTKEPAPT-TPKEPAPTAPK 488

Query:  490 KPSPQT-RLPAMIT-KTPAQLRSVA---TILK----TLCLASPTVANVKAPPQVAVAAGT 540
            KP+P T + PA  T K PA     +       T K      T    ++PT     AP     A  T
Sbjct:  489 KPAPTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTT 548

Query:  541 PNT-SGSIHENP----PKAKATVNVKQAAKVV-KASSPSYLAEGKIRCLAQPHPGTGVPR 594
            P   S + + P      PK  A   K+ A     K +P+ E     +P P    P+
Sbjct:  549 PKEPSPTTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTTKKPAPTA--PK 606

Query:  595 AAAELPLEAEKIKTGTQKQAKTDMAFKTSVAVEMAGAPSWTK-VAEEGDKPPHVYVPVDM 653
              A  P   ++    T K+         K +      AP+ + +A    + P    P +
Sbjct:  607 EPA--PTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEP 664

Query:  654 AVTLPRGQLAAPLTNASSQRHP-PCLSQRPLAAPLTKASSQGHLPTELTKTPSLAHLDTC 712
            A T P+    AAP T       + P P   + P  AP T           P E   T      T
Sbjct:  665 APTTPKA--AAPNT----PKEPAPTTPKEP--APTTPKEPAPTTPKETAPTTPKGTAPTT 716

Query:  713 LSK 715
            L +
Sbjct:  717 LKE 719

 Score = 108 (16.2 bits), Expect = 4.3e-02, Sum P(2) = 4.3e-02
 Identities = 60/214 (28%), Positives = 85/214 (39%)

Query:  265 TIRSTCLVHIEGDSVKTKRVSAR-TNKA--RAPETP-LSRRYDQAVTRPSRAQTQGPVKA 320
            T +     +H    D    T  +SA  T KA   +P+ P +          A T+P     T
Sbjct:  862 TTKEPTTIHKSPDE-STPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTT 920

Query:  321 ETP--KAPFQICPGPMITK-TLLQTYPVVSVTLPQTYPASTMTTTPPKTSPVPKVTIIKT 377
            E      P      P +TK T  T           +  T    T TTT   T+P  K+T  +KT
Sbjct:  921 ERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQD-TTPF-KITTLKT 978

Query:  378 PAQMYPGPTVTK---TAPHTCPMPTMT-KIQVHPTASRTGTPRQTCPATITAKNRPQVSL 433
               + P  T TK   T        P T K +   T S+ TP+    P    A  +P   +
Sbjct:  979 TT-LAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTK--APKKPTSTK 1035

Query:  434 LASIMKSL--PQVCPGPA-MAKTPPQMHPVTTPAKNPLQT 470
                 M  +  P+  P P  M  T P+++P +  A+  LQT
Sbjct: 1036 KPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQT 1075

 Score = 56 (8.4 bits), Expect = 3.1e-12, Sum P(2) = 3.1e-12
```

```
 Identities = 17/60 (28%), Positives = 22/60 (36%)

Query:    22 TVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKS-KEHLPQQPAEGKTASRRVP 80
             T EP T       P  P  PS      E AP  P+      + K+  P  P E    + + P
Sbjct:   533 TTKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEP 592

 Score = 52 (7.8 bits), Expect = 9.6e-16, Sum P(2) = 9.6e-16
 Identities = 17/59 (28%), Positives = 22/59 (37%)

Query:    22 TVHEPV-VTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKSKEHLPQQPAE-GKTASRR 78
             T  EP    T       P  P  P+       E  P  P+      +KE  P  P E    TA ++
Sbjct:   431 TPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKK 489

 Score = 51 (7.7 bits), Expect = 1.2e-15, Sum P(2) = 1.2e-15
 Identities = 15/51 (29%), Positives = 19/51 (37%)

Query:    22 TVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKS-KEHLPQQPAE 71
             T  EP  T       P  P  P+       + AP  P+      + KE  P  P E
Sbjct:   416 TTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTTPKE 466

 Score = 47 (7.1 bits), Expect = 3.2e-15, Sum P(2) = 3.2e-15
 Identities = 12/41 (29%), Positives = 17/41 (41%)

Query:    36 PAPAHPSLLDKMEKAPPQPQHEGLKSKEHLPQQPAEGKTAS 76
             P P  P    +       P +P     +KS      P++PA    T S
Sbjct:   350 PTPTTPK--EPASTTPKEPTPTTIKSAPTTPKEPAPTTTKS 388

 Score = 47 (7.1 bits), Expect = 3.2e-15, Sum P(2) = 3.2e-15
 Identities = 15/57 (26%), Positives = 19/57 (33%)

Query:    22 TVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEG-LKSKEHLPQQPAEGKTASR 77
             T  EP  T       P  P  P+       E AP  P+      +KE  P          T  +
Sbjct:   377 TPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPK 433

 Score = 46 (6.9 bits), Expect = 4.0e-15, Sum P(2) = 4.0e-15
 Identities = 16/58 (27%), Positives = 22/58 (37%)

Query:    20 LATVHEPVVT---QWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKSKEHLPQQPAEGKT 74
             L T  EP  T   + A   P   P+ +         P +P      KS      P++PA    T
Sbjct:   344 LTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTT 401

 Score = 42 (6.3 bits), Expect = 1.0e-14, Sum P(2) = 1.0e-14
 Identities = 15/60 (25%), Positives = 21/60 (35%)

Query:    22 TVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKS-KEHLPQQPAEGKTASRRVP 80
             T  EP  T       P  P  P+       + AP  P+      + KE  P      E    + + P
Sbjct:   463 TPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSPTTPKEP 522

 Score = 39 (5.9 bits), Expect = 2.1e-14, Sum P(2) = 2.1e-14
 Identities = 15/55 (27%), Positives = 20/55 (36%)

Query:    22 TVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLKSKEHLPQQPAEGKTAS 76
             T  EP  T       P  PA  +  +       P +P      KS      ++PA    T S
Sbjct:   494 TPKEPAPTT----PKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTKS 544
```

Pedant information for DKFZphtes3_4o19, frame 2

Report for DKFZphtes3_4o19.2

[LENGTH]	1180
[MW]	127693.40
[pI]	10.25
[HOMOL]	SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2). 1e-08
[FUNCAT]	98 classification not yet clear-cut [S. cerevisiae, YJR151c] 6e-06
[FUNCAT]	30.01 organization of cell wall [S. cerevisiae, YIR019c] 6e-06
[FUNCAT]	30.90 extracellular/secretion proteins [S. cerevisiae, YIR019c] 6e-06
[FUNCAT]	01.05.01 carbohydrate utilization [S. cerevisiae, YIR019c] 6e-06
[BLOCKS]	BL00412B Neuromodulin (GAP-43) proteins
[PROSITE]	CYTOCHROME_C 1
[PROSITE]	MYRISTYL 12
[PROSITE]	CAMP_PHOSPHO_SITE 1
[PROSITE]	CK2_PHOSPHO_SITE 8
[PROSITE]	PKC_PHOSPHO_SITE 25
[PROSITE]	ASN_GLYCOSYLATION 2
[KW]	Alpha_Beta
[KW]	LOW_COMPLEXITY 5.00 %

```
SEQ   MTLQGRADLSGNQGNAAGRLATVHEPVVTQWAVHPPAPAHPSLLDKMEKAPPQPQHEGLK
SEG   ............................................................
PRD   cccccceeeccccccceeeeeeeeceeeeeeeeccccccccceeeecccccccccccccc

SEQ   SKEHLPQQPAEGKTASRRVPRLRAVVESQAFKNILVDEMDMMHARAATLIQANWRGYWLR
SEG   ............................................................
PRD   ccccccccccccccccchhhhhhhhhhhhhhheeehhhhhhhhhhhhhhhhhhccchhhh

SEQ   QKLISQMMAAKAIQEAWRRFNKRHILHSSKSLVKKTRAEEGDIPYHAPQQVRFQHPEENR
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhheeeecccchhhhhhhccccccccccceeeecccccce

SEQ   LLSPPIMVNKETQFPSCDNLVLCRPQSSPLLQPPAAQGTPEPCVQGPHAARVRGLAFLPH
SEG   ............................................................
PRD   eeccceeeecccccccccceeeeccccccccccccccccccccccccceeeeeeeeeccc

SEQ   QTVTIRFPCPVSLDAKCQPCLLTRTIRSTCLVHIEGDSVKTKRVSARTNKARAPETPLSR
SEG   ............................................................
PRD   eeeeeecccccccccccccccccccccccceeeeeccccccccceeeeecccccccccccc

SEQ   RYDQAVTRPSRAQTQGPVKAETPKAPFQICPGPMITKTLLQTYPVVSVTLPQTYPASTMT
SEG   ........................................................xxxx
PRD   ccceeeeecccccccccceeecccccccccccccccccccccccccccccccccccccccc

SEQ   TTPPKTSPVPKVTIIKTPAQMYPGPTVTKTAPHTCPMPTMTKIQVHPTASRTGTPRQTCP
SEG   xxxxxxxxxxxxx...............................................
PRD   cccccccccccceeecccccccccccccccccccccccccccccceeeccccccccccccc

SEQ   ATITAKNRPQVSLLASIMKSLPQVCPGPAMAKTPPQMHPVTTPAKNPLQTCLSATMSKTS
SEG   ............................................................
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   SQRSPVGVTKPSPQTRLPAMITKTPAQLRSVATILKTLCLASPTVANVKAPPQVAVAAGT
SEG   ............................................................
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   PNTSGSIHENPPKAKATVNVKQAAKVVKASSPSYLAEGKIRCLAQPHPGTGVPRAAAELP
SEG   ............xxxxxxxxxxxxxxxxx.........................xxxxxx
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   LEAEKIKTGTQKQAKTDMAFKTSVAVEMAGAPSWTKVAEEGDKPPHVYVPVDMAVTLPRG
SEG   xxxx........................................................
PRD   ccccccccccccccccccccccccccccccccccceeeeccccccccceeecccccccccc

SEQ   QLAAPLTNASSQRHPPCLSQRPLAAPLTKASSQGHLPTELTKTPSLAHLDTCLSKMHSQT
SEG   ............................................................
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   HLATGAVKVQSQAPLATCLTKTQSRGQPITDITTCLIPAHQAADLSSNTHSQVLLTGSKV
SEG   ............................................................
PRD   ccccceeeeeccccceeeecccccccccccccccccccccccccccccccceeeeeccccc

SEQ   SNHACQRLGGLSAPPWAKPEDRQTQPQPHGHVPGKTTQGGPCPAACEVQGMLVPPMAPTG
SEG   ............................................................
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   HSTCNVESWGDNGATRAQPSMPGQAVPCQEDTGPADAGVVGGQSWNRAWEPARGAASWDT
SEG   ............................................................
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   WRNKAVVPPRRSGEPMVSMQAAEEIRILAVITIQAGVRGYLARRRIRLWHRGAMVIQATW
SEG   ............................................................
PRD   ccceeecccccccccchhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   RGYRVRRNLAHLCRATTTIQSAWRGYSTRRDQARHWQMLHPVTWVELGSRAGVMSDRSWF
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeeecccchhhhhhhhhh

SEQ   QDGRARTVSDHRCFQSCQAHACSVCHSLSSRIGSPPSVVMLVGSSPRTCHTCGRTQPTRV
SEG   ............................................................
PRD   hccceeeeecccceeeecccceeeeeeeeeccccccccccceeeeeecccccccccccccee

SEQ   VQGMGQGTEGPGAVSWASAYQLAALSPRQPHRQDKAATAIQSAWRGFKIRQQMRQQQMAA
SEG   ...............................................xxxxxxxxxxxxx
PRD   eeecccccccccccchhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   KIVQATWRGHHTRSCLKNTEALLGPADPSASSRHMHWPGI
SEG   xx......................................
PRD   hhhhhhccccccchhhhhhhhccccccccccccccccccc
```

Prosite for DKFZphtes3_4o19.2

PS00001	542->546	ASN_GLYCOSYLATION	PDOC00001
PS00001	668->672	ASN_GLYCOSYLATION	PDOC00001
PS00004	282->286	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	244->247	PKC_PHOSPHO_SITE	PDOC00005
PS00005	265->268	PKC_PHOSPHO_SITE	PDOC00005
PS00005	278->281	PKC_PHOSPHO_SITE	PDOC00005
PS00005	281->284	PKC_PHOSPHO_SITE	PDOC00005
PS00005	285->288	PKC_PHOSPHO_SITE	PDOC00005
PS00005	288->291	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00005	424->427	PKC_PHOSPHO_SITE	PDOC00005
PS00005	481->484	PKC_PHOSPHO_SITE	PDOC00005
PS00005	610->613	PKC_PHOSPHO_SITE	PDOC00005
PS00005	671->674	PKC_PHOSPHO_SITE	PDOC00005
PS00005	679->682	PKC_PHOSPHO_SITE	PDOC00005
PS00005	900->903	PKC_PHOSPHO_SITE	PDOC00005
PS00005	959->962	PKC_PHOSPHO_SITE	PDOC00005
PS00005	987->990	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1015->1018	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1049->1052	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1065->1068	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1106->1109	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1146->1149	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1171->1174	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	546->550	CK2_PHOSPHO_SITE	PDOC00006
PS00006	848->852	CK2_PHOSPHO_SITE	PDOC00006
PS00006	988->992	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1003->1007	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1027->1031	CK2_PHOSPHO_SITE	PDOC00006
PS00008	11->17	MYRISTYL	PDOC00008
PS00008	14->20	MYRISTYL	PDOC00008
PS00008	539->545	MYRISTYL	PDOC00008
PS00008	591->597	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00008	777->783	MYRISTYL	PDOC00008
PS00008	853->859	MYRISTYL	PDOC00008
PS00008	878->884	MYRISTYL	PDOC00008
PS00008	882->888	MYRISTYL	PDOC00008
PS00008	1008->1014	MYRISTYL	PDOC00008
PS00008	1053->1059	MYRISTYL	PDOC00008
PS00008	1083->1089	MYRISTYL	PDOC00008
PS00190	1042->1048	CYTOCHROME_C	PDOC00169

(No Pfam data available for DKFZphtes3_4o19.2)

DKFZphtes3_50j4
----

group: testes derived

DKFZphtes3_50j4 encodes a novel 187 amino acid protein proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown, prolin ritch protein

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1186 bp
Poly A stretch at pos. 1176, polyadenylation signal at pos. 1126

```
   1 CACTGGGCGT CTGAAGCTCA GAGCTCACCC CTGAGATGGG CTCTCCTAGG
  51 CCTCCTGGGA TGAGGGAGCC ACCAGGACCC AGTGCTGTGA TGCCTGCTCT
 101 TCCCTCTACC AGCACCTGCC CGCCCAGAGA CCAGGGCACC CCTGAAGTCC
 151 AGCCCACCCC TGCAAAGGAC ACATGGAAGG GCAAGCGGCC TCGATCCCAG
 201 CAGGAGAACC CAGAGAGCCA GCCTCAGAAG AGGCCACGCC CCTCAGCCAA
 251 GCCCTCCGTC GTAGCTGAGG TCAAGGGCAG CGTCTCGGCC AGCGAACAGG
 301 GCACCTTGAA TCCCACGGCT CAAGACCCCT TCCAGCTCTC CGCTCCTGGC
 351 GTCTCCTTGA AGGAGGCTGC AAATGTTGTG GTCAAGTGCC TCACCCCTTT
 401 CTACAAGGAG GGCAAGTTTG CTTCCAAGGA GTTGTTTAAA GGCTTTGCCC
 451 GCCACCTCTC ACACTTGCTG ACTCAGAAGA CCTCTCCTGG AAGGAGCGTG
 501 AAAGAAGAGG CCCAGAACCT CATCAGGCAC TTCTTCCATG GCCGGGCCCG
 551 GTGCGAGAGC GAAGCTGACT GGCATGGCCT GTGTGGCCCC CAGAGATGAC
 601 CAACTGCTGG CTGGGCAGGG CCCGCGTCCT CCCCCAGATT CTAGCATGGG
 651 TCATCCTGGG CCTCACCTGC TGATGCCAGG GCCATCGTCT TTTCTCAGTC
 701 CTTCTCCTTT CCAACCATAC TTGGCTTTGG GGATGACCCC AGACACCCCC
 751 TGAATCCAGG TCAGAGGTCA GCCCACCTTT CTTTCTGCTT GCAAAGCCTA
 801 TAGACCCTTC TCAGAGCGGT CCTCATGGCT GGGTTTTCTG GGACACATGT
 851 CGAGGACAGA AGGTGGAGGG TGGTGGAGCT GCTGCTGGAA GAAGGGGAAG
 901 GAAGAGTGGC CCCTCCCCGA GTTCTAAGTC AGGATGAGGC CCACCTGTCC
 951 AAGGTATCGG AACCTACCCA GGGGACCCTC AGATCCTCCA CCCACTCCCC
1001 CATCCATTAC GATGCCAGCT TCCAGCCTTG CCCAGGTCAG AGCTGTGGCA
1051 GAGGAGAGGC AGCCAGGCCC TGTTCCTGCT CAGCTCCTGC TCAGGAAGGC
1101 CAGGCCTGAC AGATGTTTGG GAGAGGAATA AAGTTGTGTT GTTGTGGGGC
1151 ATGCAGGCGT GCACACAGCC CTTTTCAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 36 bp to 596 bp; peptide length: 187
Category: putative protein

```
  1 MGSPRPPGMR EPPGPSAVMP ALPSTSTCPP RDQGTPEVQP TPAKDTWKGK
 51 RPRSQQENPE SQPQKRPRPS AKPSVVAEVK GSVSASEQGT LNPTAQDPFQ
101 LSAPGVSLKE AANVVVKCLT PFYKEGKFAS KELFKGFARH LSHLLTQKTS
151 PGRSVKEEAQ NLIRHFFHGR ARCESEADWH GLCGPQR
```

BLASTP hits

Entry MMU92455_1 from database TREMBL:

product: "WW domain binding protein 7"; Mus musculus WW domain binding protein 7 mRNA, partial cds.
Score = 134, P = 6.9e-08, identities = 45/125, positives = 56/125

Alert BLASTP hits for DKFZphtes3_50j4, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3_50j4, frame 3

Report for DKFZphtes3_50j4.3

[LENGTH]	187	
[MW]	20353.06	
[pI]	9.76	
[PROSITE]	MYRISTYL	1
[PROSITE]	AMIDATION	1
[PROSITE]	CK2_PHOSPHO_SITE	6
[PROSITE]	PKC_PHOSPHO_SITE	6
[KW]	All_Alpha	
[KW]	LOW_COMPLEXITY	8.56 %

```
SEQ  MGSPRPPGMREPPGPSAVMPALPSTSTCPPRDQGTPEVQPTPAKDTWKGKRPRSQQENPE
SEG  xxxxxxxxxxxxxxxx............................................
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SQPQKRPRPSAKPSVVAEVKGSVSASEQGTLNPTAQDPFQLSAPGVSLKEAANVVVKCLT
SEG  ............................................................
PRD  cccccccccccccchhhhhcccccccccccccccccccccccccccccchhhhhhheeecc

SEQ  PFYKEGKFASKELFKGFARHLSHLLTQKTSPGRSVKEEAQNLIRHFFHGRARCESEADWH
SEG  ............................................................
PRD  cccccccchhhhhhhhhhhhhhhhhheeeccccccchhhhhhhhhhhhhhcccchhhhhhhhh

SEQ  GLCGPQR
SEG  .......
PRD  ccccccc
```

Prosite for DKFZphtes3_50j4.3

PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	70->73	PKC_PHOSPHO_SITE	PDOC00005
PS00005	107->110	PKC_PHOSPHO_SITE	PDOC00005
PS00005	146->149	PKC_PHOSPHO_SITE	PDOC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	84->88	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	175->179	CK2_PHOSPHO_SITE	PDOC00006
PS00008	81->87	MYRISTYL	PDOC00008
PS00009	48->52	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_50j4.3)

DKFZphtes3_50n06
----

group: testes derived

DKFZphtes3_50n06 encodes a novel 186 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1095 bp
Poly A stretch at pos. 1085, polyadenylation signal at pos. 1061

```
   1 CAAGACCCTC GGAGCCAAGA AACAACACTG AGTTCCAGAT TTCGGAAGGT
  51 TCACGAGTGT TGCCGACACG CCCTCCCAAC TGCAGACATC CTCCCTGGAG
 101 GACCTGCTGT GCTCACATGC CCCCCTGTCC AGCGAGGACG ACACCTCCCC
 151 GGGCTGTGCA GCCCCCTCCC AGGCACCCTT CAAGGCCTTC CTCAGTCCCC
 201 CAGAGCCACA TAGCCACCGA GGCACCGACA GGAAGCTGTC CCCGCTCCTG
 251 AGCCCCTTGC AAGACTCACT GGTGGACAAG ACCCTGCTGG AGCCCAGGGA
 301 GATGGTCCGG CCTAAGAAGG TGTGTTTCTC GGAGAGCAGC CTGCCCACCG
 351 GGGACAGGAC CAGGAGGAGC TACTACCTCA ATGAGATCCA GAGCTTCGCG
 401 GGCGCCGAGA AGGACGCGCG CGTGGTGGGC GAGATCGCCT TCCAGCTGGA
 451 CCGCCGCATC CTGGCCTACG TGTTCCCGGG CGTGACGCGG CTCTACGGCT
 501 TCACGGTGGC CAACATCCCC GAGAAGATCG AGCAGACCTC CACCAAGTCT
 551 CTGGACGGCT CCGTGGACGA GAGGAAGCTG CGCGAGCTGA CGCAGCGCTA
 601 CCTGGCCCTG AGCGCGCGCC TGGAGAAGCT GGGCTACAGC CGCGACGTGC
 651 ACCCGGCGTT CAGCGAGTTC CTCATCAACA CCTACGGAAT CCTGAAGCAG
 701 CGGCCCGACC TGCGCGCCAA CCCCCTGCAC AGCAGCCCGG CCGCGCTGCG
 751 CAAGCTGGTC ATCGACGTGG TGCCCCCCAA GTTCCTGGGC GACTCGCTGC
 801 TGCTGCTCAA CTGCCTGTGC GAGCTCTCCA AGGAGGACGG CAAGCCCCTC
 851 TTCGCCTGGT GAGCCGCCCC GCGCCCGCCG CCTTGCCTGC AGTAAACGCG
 901 TTTGTTCCAA CCCGGGGCCG CGGTGCCTCC TGCGCGTCCC CCCGGAGGGG
 951 AAAGGGCCGC GTCCCCCGCG CGCGAGGCCA GAGAAGGCCC CGCTCCCACC
1001 GGTGCTGGGC CCCGACCGCA GCCCGCCGCT GCCCGCACCT GCGGAGTGCT
1051 TCTCACCCCT CATTAAAATC ATCCGTTTGC TTGTCAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 302 bp to 859 bp; peptide length: 186
Category: putative protein
Classification: no clue

```
  1 MVRPKKVCFS ESSLPTGDRT RRSYYLNEIQ SFAGAEKDAR VVGEIAFQLD
 51 RRILAYVFPG VTRLYGFTVA NIPEKIEQTS TKSLDGSVDE RKLRELTQRY
101 LALSARLEKL GYSRDVHPAF SEFLINTYGI LKQRPDLRAN PLHSSPAALR
151 KLVIDVVPPK FLGDSLLLLN CLCELSKEDG KPLFAW
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_50n06, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_50n06, frame 2

Report for DKFZphtes3_50n06.2

```
[LENGTH]        186
[MW]            21049.39
[pI]            9.28
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY     5.38 %

SEQ   MVRPKKVCFSESSLPTGDRTRRSYYLNEIQSFAGAEKDARVVGEIAFQLDRRILAYVFPG
SEG   ............................................................
PRD   cccccceeecccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccc

SEQ   VTRLYGFTVANIPEKIEQTSTKSLDGSVDERKLRELTQRYLALSARLEKLGYSRDVHPAF
SEG   ............................................................
PRD   ceeeeeeeeeeccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccccccch

SEQ   SEFLINTYGILKQRPDLRANPLHSSPAALRKLVIDVVPPKFLGDSLLLLNCLCELSKEDG
SEG   .............................................xxxxxxxxxx.....
PRD   hhhhhhcceeeccccccccccccccchhhhhhhhhhcccccccccchhhhhhhhhhhcccc

SEQ   KPLFAW
SEG   ......
PRD   cccccc
```

(No Prosite data available for DKFZphtes3_50n06.2)

(No Pfam data available for DKFZphtes3_50n06.2)

DKFZphtes3_50n23

group: testes derived

DKFZphtes3_50n23 encodes a novel 499 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

2 EST hits
(from other testis librarys) testis specific cDNA?

Sequenced by DKFZ

Locus: unknown

Insert length: 1907 bp
Poly A stretch at pos. 1897, polyadenylation signal at pos. 1872

```
   1 GGGCACCAGC CACTTTCCAC CATGACTGTG CGCTCGAGGG TCGCAGATGT
  51 GTTCGGCAGC AAGGACACTG AGAGCCTTGA GCCTGTGCTT TTACCCTTAG
 101 TAGATCGCAG GTTTCCTAAG AAATGGGAAA GACCGGTGGC AGAAAGCTTA
 151 GGCCACAAAG ACAAAGACCA GGAGGACTAC TTCCAGAAGG GAGGACTCCA
 201 AATTAAGTTC CACTGTAGCA AGCAGCTGTC TCTAGAGAGC TCCAGGCAGG
 251 TGACCTCTGA GAGCCAAGAG GAGCCCTGGG AGGAGGAATT CGGCCGGGAG
 301 ATGCGGAGGC AGCTGTGGCT GGAGGAGGAG GAGATGTGGC AGCAGCGGCA
 351 GAAGAAGTGG GCCCTGCTGG AGCAGGAGCA TCAGGAGAAG CTGCGGCAGT
 401 GGAATCTGGA AGACCTGGCC AGGGAGCAAC AGCGGAGATG GGTCCAGCTA
 451 GAAAAGGAGC AGGAGAGCCC ACGGAGAGAG CCAGAGCAGC TAGGGGAGGA
 501 TGTGGAGAGG AGGATCTTCA CACCCACCAG TCGATGGAGG GACTTGGAGA
 551 AGGCAGAGCT ATCATTAGTG CCTGCCCCAA GCCGGACCCA ATCTGCTCAC
 601 CAAAGCAGGA GGCCACACTT GCCCATGTCT CCTAGTACCC AGCAGCCTGC
 651 CCTGGGAAAG CAGAGACCTA TGAGTTCAGT GGAGTTTACC TACAGACCAC
 701 GGACCCGCCG AGTTCCCACA AAGCCCAAGA AATCTGCCTC CTTTCCTGTC
 751 ACTGGGACAT CCATCCGAAG GCTGACCTGG CCCTCTTTGC AGATATCCCC
 801 TGCAAATATT AAGAAGAAGG TGTACCACAT GGACATGGAG GCCCAGAGGA
 851 AGAACCTGCA GCTCCTGAGT GAGGAGTCTG AGTTGAGGCT GCCCCACTAC
 901 CTGCGCAGCA AAGCACTGGA GCTCACCACC ACCACCATGG AGCTGGGCGC
 951 GCTCAGGCTG CAGTACCTGT GCCATAAGTA CATCTTCTAT AGACGCCTCC
1001 AGAGCCTCCG GCAAGAAGCG ATCAACCATG TACAAATCAT GAAAGAAACG
1051 GAGGCTTCCT ACAAGGCCCA GAACCTCTAC ATCTTCCTGG AAAACATTGA
1101 CCGCCTGCAG AGTCTCAGGC TGCAGGCCTG GACGGACAAG CAGAAGGGGC
1151 TGGAGGAGAA GCACCGAGAG TGCCTGAGCA GCATGGTGAC CATGTTCCCC
1201 AAGCTCCAGC TGGAGTGGAA CGTTCACCTG AACATCCCTG AGGTCACCTC
1251 GCCAAAGCCA AAGAAATGCA AGTTGCCTGC AGCCTCACCC CGGCACATCC
1301 GCCCCAGTGG CCCCACCTAC AAGCAGCCCT TTCTGTCTAG GCACCGGGCA
1351 TGTGTGCCCC TGCAGATGGC CCGCCAACAG GGGAAGCAGA TGGAGGCTGT
1401 CTGGAAGACC GAGGTGGCCT CCTCCAGTTA CGCAATAGAA AAAAAGACCC
1451 CTGCCAGCCT TCCCCGGGAC CAGCTGAGGG GACACCCAGA TATTCCCCGG
1501 CTGTTGACAC TGGACGTGTA GTCCTCCTGC CACAAAAGCC TGAACTTCCT
1551 GAAGGCCCAG TAAGCGCCTC AGCGAACCAA AGGAAGGAAT GCCAGGAACC
1601 TACAAATGAA TCCGCTTAGC TTGTTCAAAA AAAGTCAAGC GAGTCACTCC
1651 CTGGAACCCA AATAAGCCAG AAGGATCAAG ACAGCCCCAG TCTCCACTGC
1701 ATCCCTCAGC CAGTGATTCT CAACCTTCTG AGGGACGGAA ACCCACAGAG
1751 AACTTGGTCA AAATGCAGGT TCCCAGCTGG TGCTTTTAAA GAAACCCTCT
1801 GGGGGTTGCT GAGTACTCCT AGAACTTTGA GAAACACTGC TTCCCTCCTG
1851 CAGTCCCCAA ACTCTACATT TTAATAAAAT AGAGGTTGGT TTATTTTAAA
1901 AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

```
                    Peptide information for frame 1
                    -------------------------------

ORF from 22 bp to 1518 bp; peptide length: 499
Category: similarity to known protein
Classification: no clue


     1 MTVRSRVADV FGSKDTESLE PVLLPLVDRR FPKKWERPVA ESLGHKDKDQ
    51 EDYFQKGGLQ IKFHCSKQLS LESSRQVTSE SQEEPWEEEF GREMRRQLWL
   101 EEEEMWQQRQ KKWALLEQEH QEKLRQWNLE DLAREQQRRW VQLEKEQESP
   151 RREPEQLGED VERRIFTPTS RWRDLEKAEL SLVPAPSRTQ SAHQSRRPHL
   201 PMSPSTQQPA LGKQRPMSSV EFTYRPRTRR VPTKPKKSAS FPVTGTSIRR
   251 LTWPSLQISP ANIKKKVYHM DMEAQRKNLQ LLSEESELRL PHYLRSKALE
   301 LTTTTMELGA LRLQYLCHKY IFYRRLQSLR QEAINHVQIM KETEASYKAQ
   351 NLYIFLENID RLQSLRLQAW TDKQKGLEEK HRECLSSMVT MFPKLQLEWN
   401 VHLNIPEVTS PKPKKCKLPA ASPRHIRPSG PTYKQPFLSR HRACVPLQMA
   451 RQQGKQMEAV WKTEVASSSY AIEKKTPASL PRDQLRGHPD IPRLLTLDV

                                 BLASTP hits

No BLASTP hits available

            Alert BLASTP hits for DKFZphtes3_50n23, frame 1

PIR:S28589 trichohyalin - rabbit, N = 1, Score = 134, P = 5.3e-05

TREMBLNEW:AF132479_1 product: "Ese2L protein";      Mus musculus Ese2L
protein mRNA, complete cds., N = 1, Score = 130, P = 0.00017


>PIR:S28589 trichohyalin - rabbit
            Length = 1,407

  HSPs:

 Score = 134 (20.1 bits), Expect = 5.3e-05, P = 5.3e-05
 Identities = 88/354 (24%), Positives = 154/354 (43%)

Query:     29 RRFPKKWERPVAESLGHKDKDQEDYFQKGGLQIK-FHCSKQLSLESSRQVTSESQEEPWE 87
              R++  K +R   + L  +  ++E   ++ G   + F   +QL     +++  E +EE  +
Sbjct:    165 RQYRDKEQRLQRQELEERRAEEEQLRRRKGRDAEEFIEEEQLRRREQQELKRELREEEQQ 224

Query:     88 EEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRRWVQLEKEQ 147
                  RE   +  L+EEE    RQ++W   E   Q++LR+  LE++ RE+++R  Q E+ +
Sbjct:    225 RRERREQHERA-LQEEEEQLLRQRRWRE-EPREQQQLRR-ELEEI-REREQRLEQEERRE 280

Query:    148 ESPRREPEQLGEDVERRIFTPTSRWRDLEKAELSLVPAPSRTQSAHQSRRPHLPMSPSTQ 207
              +  RRE ++L E  ERR         ++ + E  L     R Q   Q  R    +    +
Sbjct:    281 QQLRRE-QRL-EQEERREQQLRRELEEIREREQRLEQEERREQRLEQEERREQQLKRELE 338

Query:    208 QPALGKQRPMSSVEFTYRPRTRRVPTKPKKSASFPVTGTSIRRLTWPSLQISPANIKK-K 266
              +    +QR    +E   R R + +  + ++ A     G S+ R  W     S A  ++ K
Sbjct:    339 EIREREQR----LEQEER-REQLLAEEVREQAR--ERGESLTR-RWQRQLESEAGARQSK 390

Query:    267 VYHMDMEAQRKNLQLLSEESELRLPHYLRSKALELTTTTM------ELGALRLQYLCHKY 320
              VY      +R+  Q L ++ E R     R + LE            E    R Q L  +
Sbjct:    391 VYS---RPRRQEEQSLRQDQERR-QRERERELEEQARRQQQWQAEEESERRRQRLSARP 446

Query:    321 IFYRRLQSLRQEAINHVQIMKETEASYKAQNLYI-FLENIDRLQSL-RLQAWTDKQKGLE 378
                  R Q   +E     Q  +E E   + +   + FLE  ++LQ   R Q   ++    E
Sbjct:    447 SLRER-QLRAEERQEQEQRFREEEEQRRERRQELQFLEEEEQLQRRERAQQLQEEDSFQE 505

Query:    379 EKHR 382
              ++ R
Sbjct:    506 DRER 509

 Score = 119 (17.9 bits), Expect = 2.2e-03, P = 2.2e-03
 Identities = 79/357 (22%), Positives = 150/357 (42%)

Query:     33 KKWERPVAESLGHKDKDQEDYFQKGGLQIKFHCSKQLSLESSRQVTSESQEEPWEEEFGR 92
              ++ E+ + +    K +++E   Q+   + +    +Q      R+   + + + EE+F  +
Sbjct:    990 RREEQELRQERDRKFREEEQLLQE---REEERLRRQERDRKFREEERQLRRQELEEQFRQ 1046

Query:     93 EMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRRWVQLEKEQESPRR 152
              E  R+  LEE+ + Q++++K  L  QE   K R+   E+  R +Q R  QL +E++   R
Sbjct:   1047 ERDRKFRLEEQ-IRQEKEEK-QLRRQERDRKFRE---EEQQRRRQEREQQLRRERDRKFR 1101

Query:    153 EPEQLGEDVERRIFTPTSRWRDLEKAELSLVPAPSRTQSAHQSR--RPHLPMSPSTQQPA 210
```

```
            E EQL ++ E        R R L + E  L+     +    + R  R    +   +++
Sbjct: 1102 EEEQLLQEREEERLRRQERARKLREEE-QLLRREEQLLRQERDRKFREEEQLLQESEEER 1160

Query:  211 LGKQ---RPMSSVEFTYRPRTRRVPTKPKKSASFPVTGTSIRRLTWPSLQISPANIKKKV 267
            L +Q   R +   E   + R      + +++         +R+      Q     ++++
Sbjct: 1161 LRRQERERKLREEEQLLQEREEERLRRQERARKLREEEQLLRQEEQELRQERARKLREEE 1220

Query:  268 YHMDMEAQ----------RKNLQLLS-EESELRLPHYLRSKALELTTTTMELGALRLQYL 316
              + E Q          R+  QLL  EE ELR     + +  E      E   LR Q
Sbjct: 1221 QLLRQEEQELRQERDRKFREEEQLLRREEQELRRERDRKFREEEQLLQEREEERLRRQER 1280

Query:  317 CHKYIFYRRLQSLRQEAINHVQIMKETEASYKAQNLYIFLENIDRLQ-SLRLQAWTDKQK 375
              K       + L  E     ++ +E +  Y+A+  +   E   RL+  LR +    +++
Sbjct: 1281 ARK--LREEEEQLLFEEQEEQRLRQERDRRYRAEEQFAREEKSRRLERELRQEEEQRRRR 1338

Query:  376 GLEEKHRE 383
              E K RE
Sbjct: 1339 ERERKFRE 1346

 Score = 109 (16.4 bits), Expect = 1.9e-01, P = 1.7e-01
 Identities = 37/113 (32%), Positives = 60/113 (53%)

Query:   67 KQLSLESSRQVTSESQ--EEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKL 124
            +QL  E  R+   E Q  +E  EE   R+ R +   EEE++ Q+R+++  L  QE + KL
Sbjct:  764 QQLRRERDRKFREEEQLLQEREEERLRRQERERKLREEEQLLQEREEE-RLRRQERERKL 822

Query:  125 RQWNLEDLAREQQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
            R+   E L +E++   ++  +E+E   RE EQL    E+ +     R R L + E
Sbjct:  823 REE--EQLLQEREEERLR-RQERERKLREEEQLLRQEEQEL--RQERARKLREEE 872

 Score = 107 (16.1 bits), Expect = 3.0e-01, P = 2.6e-01
 Identities = 35/109 (32%), Positives = 61/109 (55%)

Query:   71 LESSRQVTSESQEEPWE-EEFGREMRRQL---WLEEEEMWQQRQKKWALLEQEHQEKLRQ 126
            L    Q+  ES+EE    +E  +++RR+    + EEE++ Q+R+++  L  QE + KLR+
Sbjct:  742 LREEEQLLQESEEERLRRQEREQQLRRERDRKFREEEQLLQEREEE-RLRRQERERKLRE 800

Query:  127 WNLEDLAREQQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
               E L +E++   ++  +E+E   RE EQL ++ E        R R L + E
Sbjct:  801 E--EQLLQEREEERLR-RQERERKLREEEQLLQEREEERLRRQERERKLREEE 850

 Score = 104 (15.6 bits), Expect = 9.4e-02, P = 9.0e-02
 Identities = 84/339 (24%), Positives = 149/339 (43%)

Query:   67 KQLSLESSRQVTSESQEEPWEEEFGREMRRQL-WLEEEEMWQQRQKKWALLEQE--HQEK 123
            +QL  E  ++     +EE   EE  RE R++L +LEEEE  Q+R++   L E++   +++
Sbjct:  451 RQLRAEERQEQEQRFREE---EEQRRERRQELQFLEEEEQLQRRERAQQLQEEDSFQEDR 507

Query:  124 LRQWNLEDLAREQQRRWVQLEKEQESPRR----EP---EQLGEDVE-RRIFTPTSRWRDL 175
             R+   ++    Q  RW QL++E + R     +P   EQL E+ E +R      R R+
Sbjct:  508 ERRRRQQEQRPGQTWRW-QLQEEAQRRRHTLYAKPGQQEQLREEEELQREKRRQEREREY 566

Query:  176 EKAELSLVPAPSRTQSAHQSRRPHLPMSPSTQQPALGKQRPMSSVEFTYRPRT----RRV 231
             + E  L       +   +  R +  +    Q+  L + R +   E   + R     RR
Sbjct:  567 REEE-KLQREEDEKRRRQERERQYRELEELRQEEQL-RDRKLREEEQLLQEREEERLRRQ 624

Query:  232 PTKPK---KSASFPVTGTSIRRLTWPSLQISPANIKKKVYHMDMEAQRK---NLQLLSEE 285
              + K   +          +R+     L+     ++++   +  E +RK     QLL E
Sbjct:  625 ERERKLREEEQLLRQEEQELRQERERKLREEEQLLRREEQELRQERERKLREEEQLLQER 684

Query:  286 SELRLPHYLRSKALE-----LTTTTMELGALRLQYLCHKYIFYRRL-QSLRQEAINHV-- 337
             E RL    R++ L      L     EL   R + L  +    RR  Q LRQE    +
Sbjct:  685 EEERLRRQERARKLREEEQLLRQEEQELRQERERKLREEEQLLRREEQLLRQERDRKLRE 744

Query:  338 --QIMKETEASYKAQNLYIFLENIDRLQSLRLQAWTDKQKGLEEKHRECL 385
              Q+++E+E      +     E   +L+  R + + ++++ L+E+  E L
Sbjct:  745 EEQLLQESEEERLRRQ-----EREQQLRRERDRKFREEEQLLQEREEERL 789

 Score = 103 (15.5 bits), Expect = 1.2e-01, P = 1.1e-01
 Identities = 42/152 (27%), Positives = 74/152 (48%)

Query:   36 ERPVAESLGHKDKDQEDYFQKGGLQIKFHCSKQLSLESSRQVTSESQEEPWEEEFG-REM 94
            ER   +    K +++E   ++   +++   +++L  E  + + E QE   E +   RE
Sbjct:  835 ERLRRQERERKLREEEQLLRQEEQELRQERARKLR-EEEQLLRQEEQELRQERDRKLREE 893

Query:   95 RRQLWLEEEEMWQQRQKKWA----LLEQEHQEKLRQWNLEDLAREQQ---RRWVQ-LEKE 146
             + L  EE+E+ Q+R +K      LL++  +E+LR+   E   RE++   RR  Q L +E
Sbjct:  894 EQLLRQEEQELRQERDRKLREEEQLLQESEEERLRRQERERKLREEEQLLRREEQELRRE 953

Query:  147 QESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
            +    RE EQL ++ E        R R L + E
```

```
Sbjct:   954 RARKLREEEQLLQEREEERLRRQERARKLREEE 986

 Score = 103 (15.5 bits), Expect = 7.8e-01, P = 5.4e-01
 Identities = 31/91 (34%), Positives = 52/91 (57%)

Query:    67 KQLSLESSRQVTSESQEEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQ 126
             ++L  E  R++  E Q    EE+  R+ R +    EEE++ Q+R+++  L  QE    KLR+
Sbjct:   642 QELRQERERKLREEEQLLRREEQELRQERERKLREEEQLLQEREEE-RLRRQERARKLRE 700

Query:   127 WNLEDLAREQQRRWVQLEKEQESPRREPEQL 157
               E L R++++    +L +E+E   RE EQL
Sbjct:   701 E--EQLLRQEEQ---ELRQERERKLREEEQL 726

 Score = 101 (15.2 bits), Expect = 2.0e-01, P = 1.8e-01
 Identities = 38/111 (34%), Positives = 57/111 (51%)

Query:    72 ESSRQVTSESQEEPWEE-EFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLE 130
             E  R++  E Q    EE E  RE  R+L  EEE++ Q+R+++  L  QE    KLR+   +
Sbjct:   931 ERERKLREEEQLLRREEQELRERARKL-REEEQLLQEREEE-RLRRQERARKLREEE-Q 987

Query:   131 DLAREQQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAELSL 182
              L RE+Q     +L +E++   RE EQL ++ E         R R   + E  L
Sbjct:   988 LLRREEQ----ELRQERDRKFREEEQLLQEREEERLRRQERDRKFREEERQL 1035

 Score = 101 (15.2 bits), Expect = 1.3e+00, P = 7.2e-01
 Identities = 33/108 (30%), Positives = 56/108 (51%)

Query:    72 ESSRQVTSESQEEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLED 131
             E  R++  E Q    EE+  R+ R +    EEE++ +Q +++   L QE   KLR+   E
Sbjct:   841 ERERKLREEEQLLRQEEQELRQERARKLREEEQLLRQEEQE---LRQERDRKLREE--EQ 895

Query:   132 LAREQQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAE 179
             L R++++    +L +E++   RE EQL ++ E           R R L + E
Sbjct:   896 LLRQEEQ---ELRQERDRKLREEEQLLQESEEERLRRQERERKLREEE 940

 Score = 99 (14.9 bits), Expect = 2.0e+00, P = 8.7e-01
 Identities = 32/97 (32%), Positives = 50/97 (51%)

Query:    72 ESSRQVTSESQEEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLED 131
             E  R+   E Q     EE     E  R     L EEE   Q +++ L  QE + KLR+   E
Sbjct:   578 EKRRRQERERQYRELEELRQEEQLRDRKLREEEQLLQEREEERLRRQERERKLREE--EQ 635

Query:   132 LAREQ-----QRRWVQLEKEQESPRREPEQLGEDVERRI 165
             L R++        Q R +L +E++  RRE ++L ++ ER++
Sbjct:   636 LLRQEEQELRQERERKLREEEQLLRREEQELRQERERKL 674

 Score = 99 (14.9 bits), Expect = 2.0e+00, P = 8.7e-01
 Identities = 34/111 (30%), Positives = 58/111 (52%)

Query:    67 KQLSLESSRQVTSESQ--EEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKL 124
             ++L  E  R++  E Q  +E  EE   R+ R +    EEE++ +Q +++   L QE + KL
Sbjct:   664 QELRQERERKLREEEQLLQEREEERLRRQERARKLREEEQLLRQEEQE---LRQERERKL 720

Query:   125 RQWNLEDLAREQQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEK 177
             R+    + L RE+Q      L +E++   RE EQL ++ E           R + L +
Sbjct:   721 REEE-QLLRREEQL----LRQERDRKLREEEQLLQESEEERLRRQEREQQLRR 768

 Score = 98 (14.7 bits), Expect = 2.6e+00, P = 9.2e-01
 Identities = 37/146 (25%), Positives = 77/146 (52%)

Query:    20 EPVLLPLVDRRFPKKWERPVAESLGHKDKDQEDYFQKGGLQIKFHCSKQLSLESSRQVTS 79
             E LL    ++    ++ ER + E        + +E+   ++     K    +QL  +   +++
Sbjct:   655 EEQLLRREEQELRQERERKLREEEQLLQEREEERLRRQERARKLREEEQLLRQEEQELRQ 714

Query:    80 ESQEEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLED-LAREQQR 138
             E + +  EEE  + +RR+  L +E   ++ ++ +++  LL++  +E+LR+ E  L RE+ R
Sbjct:   715 ERERKLREEE--QLLRREEQLLRQERDRKLREEEQLLQESEEERLRRQEREQQLRRERDR 772

Query:   139 RWVQLEKEQESPRREPEQLG-EDVERRI 165
             ++++   E+EQ  RE E+L  ++ ER++
Sbjct:   773 KF--REEEQLLQEREEERLRRQERERKL 798

 Score = 97 (14.6 bits), Expect = 3.3e+00, P = 9.6e-01
 Identities = 38/129 (29%), Positives = 63/129 (48%)

Query:    72 ESSRQVTSESQ--EEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNL 129
             E  R++  E Q  +E  EE   R+ R +    EEE++ +Q +++   L QE   KLR+
Sbjct:   817 ERERKLREEEQLLQEREEERLRRQERERKLREEEQLLRQEEQE---LRQERARKLREE-- 871

Query:   130 EDLAREQQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAELSLVPAPSRT 189
             E L R++++    +L +E++   RE EQL   E+ +     R R L + E L+
```

```
Sbjct:   872 EQLLRQEEQ---ELRQERDRKLREEEQLLRQEEQEL--RQERDRKLREEE-QLLQESEEE 925

Query:   190 QSAHQSRRPHL 200
             +   Q R   L
Sbjct:   926 RLRRQERERKL 936

 Score = 96 (14.4 bits), Expect = 4.1e+00, P = 9.8e-01
 Identities = 41/132 (31%), Positives = 69/132 (52%)

Query:    46 KDKDQEDYFQKGGLQI-KFHCSKQLSLESSRQVTSESQEEPWEEEFGREMRRQLWLEEEE 104
             +++ QE  F +   Q+ +   ++QL  E S Q   E +    E+  G+  R QL   +EE
Sbjct:   473 RERRQELQFLEEEEQLQRRERAQQLQEEDSFQEDRERRRRQQEQRPGQTWRWQL---QEE 529

Query:   105 MWQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRRWVQLEKEQESPRREPEQLGEDVERR 164
               ++R   +A   Q  QE+LR+   E+L RE++R+    E+E+E    E  Q  ED +RR
Sbjct:   530 AQRRRHTLYAKPGQ--QEQLREE--EELQREKRRQ----EREREYREEEKLQREEDEKRR 581

Query:   165 IFTPTSRWRDLEK 177
                   ++R+LE+
Sbjct:   582 RQERERQYRELEE 594

 Score = 96 (14.4 bits), Expect = 4.1e+00, P = 9.8e-01
 Identities = 35/138 (25%), Positives = 76/138 (55%)

Query:    28 DRRFPKKWERPVAESL-GHKDKDQEDYFQKGGLQIKFHCSKQLSLESSRQVTSESQEEPW 86
             +R++ +  E    E L   K +++E   Q+   +      ++  L    Q+  + ++E
Sbjct:   586 ERQYRELEELRQEEQLRDRKLREEEQLLQEREEERLRRQERERKLREEEQLLRQEEQE-L 644

Query:    87 EEEFGREMRRQLWL---EEEEMWQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRRWVQL 143
              +E  R++R +  L   EE+E+ Q+R++K   L +E Q  L++   E L R+++ R  +L
Sbjct:   645 RQERERKLREEEQLLRREEQELRQERERK---LREEEQ-LLQEREEERLRRQERAR--KL 698

Query:   144 EKEQESPRREPEQLGEDVERRI 165
              +E++  R+E ++L ++ ER++
Sbjct:   699 REEEQLLRQEEQELRQERERKL 720

 Score = 95 (14.3 bits), Expect = 5.2e+00, P = 9.9e-01
 Identities = 59/282 (20%), Positives = 121/282 (42%)

Query:    20 EPVLLPLVDRRFPKKWERPVAESLGHKDKDQEDYFQKGGLQIKFHCSKQLSLESSRQVTS 79
             E  LL   ++   ++ ER + E      + +E+  ++     K    +QL  +  +++
Sbjct:   655 EEQLLRREEQELRQERERKLREEEQLLQEREEERLRRQERARKLREEEQLLRQEEQELRQ 714

Query:    80 ESQEEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNLED-LAREQQR 138
             E + +  EEE  + +RR+  L  +E  ++ +++  LL++  +E+LR+   E  L RE+ R
Sbjct:   715 ERERKLREEE--QLLRREEQLLRQERDRKLREEEQLLQESEEERLRRQEREQQLRRERDR 772

Query:   139 RWVQLEKEQESPRREPEQLG-EDVERRIFTPTSRWRDLEKAELSLVPAPSRTQSAHQ--S 195
             ++   E+EQ    RE E+L  ++ ER++       ++ E+  L       + +   Q
Sbjct:   773 KF--REEEQLLQEREEERLRRQERERKLREEEQLLQEREEERLRRQERERKLREEEQLLQ 830

Query:   196 RRPHLPMSPSTQQPALGKQRPMSSVEFTYRPRTRRVPTKPKKSASFPVTGTSIRRLTWPS 255
              R    +    ++  L ++  +   E     R  R     ++          +R+
Sbjct:   831 EREEERLRRQERERKLREEEQLLRQE-EQELRQERARKLREEEQLLRQEEQELRQERDRK 889

Query:   256 LQISPANIKKKVYHMDMEAQRK---NLQLLSEESELRLPHYLRSKAL 299
             L+     ++++   +  E  RK     QLL E  E RL    R + L
Sbjct:   890 LREEEQLLRQEEQELRQERDRKLREEEQLLQESEEERLRRQERERKL 936

 Score = 94 (14.1 bits), Expect = 1.1e+00, P = 6.8e-01
 Identities = 35/116 (30%), Positives = 59/116 (50%)

Query:    72 ESSRQVTSESQEEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEK-------L 124
             E +R++  E Q    EE+  R+ R + + EEE++ Q+R+++  L  QE   K       L
Sbjct:   977 ERARKLREEEQLLRREEQELRQERDRKFREEEQLLQEREEE-RLRRQERDRKFREEERQL 1035

Query:   125 RQWNLEDLAREQQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAELSL 182
             R+  LE+  R+++ R  +LE EQ    +E +QL      R F    + R  ++ E  L
Sbjct:  1036 RRQELEEQFRQERDRKFRLE-EQIRQEKEEKQLRRQERDRKFREEEQQRRRQEREQQL 1092

 Score = 94 (14.1 bits), Expect = 1.1e+00, P = 6.8e-01
 Identities = 51/166 (30%), Positives = 76/166 (45%)

Query:    67 KQLSLESSRQVTSESQ--EEPWEEEFGREMR-RQLWLEEEEMWQQRQKKWALLEQEHQEK 123
             ++L  E  R+   E Q  +E  EE   R+ R R+L  EEE++  + Q++  L     QE+
Sbjct:  1250 QELRRERDRKFREEEQLLQEREEERLRRQERARKLREEEEQLLFEEQEEQRL----RQER 1305

Query:   124 LRQWNLED-LAREQQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAELSL 182
              R++  E+  ARE++ R  +LE+E    R+E EQ       R F      R  E+ E
Sbjct:  1306 DRRYRAEEQFAREEKSR--RLEREL---RQEEEQRRRRERERKFREEQLRRQQEE-EQRR 1359
```

```
Query:   183 VPAPSRTQSAHQSRRPHLPMSPSTQQPALGKQRPMSSVEFTYRPRTRRVP 232
                 R     QSRR  L    P T+Q A    R      E+    R++  P
Sbjct:  1360 RQLRERQFREDQSRRQVL--EPGTRQFARVPVRSSPLYEYIQEQRSQYRP 1407

 Score = 93 (14.0 bits), Expect = 8.3e+00, P = 1.0e+00
 Identities = 41/145 (28%), Positives = 72/145 (49%)

Query:    28 DRRFPKKWERPVAESLGHKDKDQEDYFQKGGLQIKFHCSKQLSLESSRQVTSESQEEPW- 86
             +RR   ++ ER + E    + + Q +    +    Q +        L        R      + QE+ +
Sbjct:   408 ERRQRQERERELEEQARRQQQWQAEEESERRRQ-RLSARPSLRERQLRAEERQEQEQRFR 466

Query:    87 -EEEFGREMRRQL-WLEEEEMWQQRQKKWALLEQE--HQEKLRQWNLEDLAREQQRRWVQ 142
              EEE   RE R++L +LEEEE  Q+R++    L E++    +++ R+    ++     Q  RW Q
Sbjct:   467 EEEEQRRERRQELQFLEEEEQLQRRERAQQLQEEDSFQEDRERRRRQQEQRPGQTWRW-Q 525

Query:   143 LEKEQESPRR----EP---EQLGEDVE 162
             L++E +  R      +P   EQL E+ E
Sbjct:   526 LQEEAQRRRHTLYAKPGQQEQLREEEE 552

 Score = 91 (13.7 bits), Expect = 2.4e+00, P = 9.1e-01
 Identities = 38/110 (34%), Positives = 57/110 (51%)

Query:    72 ESSRQVTSESQEEPWEE-EFGREMRRQLWLEEEEMWQQRQKKWALLEQEHQEKLRQWNL- 129
             E  R++  E Q      EE E  RE  R+L  EEE++ Q+R+++  L  QE    KLR+
Sbjct:   931 ERERKLREEEQLLRREEQELRRERARKL-REEEQLLQEREEE-RLRRQERARKLREEEQL 988

Query:   130 -----EDLAREQQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAEL 180
                  ++L +E+ R++   E+EQ    RE E+L       R F    R   L + EL
Sbjct:   989 LRREEQELRQERDRKF--REEEQLLQEREEERLRRQERDRKFREEER--QLRRQEL 1040

 Score = 89 (13.4 bits), Expect = 2.2e+00, P = 8.9e-01
 Identities = 35/138 (25%), Positives = 65/138 (47%)

Query:    82 QEEPWEEEFGREMRRQLWLEEEEM--WQQRQKKWALLEQEHQEKLRQWNLEDLAREQQRR 139
             Q     E++    E+R + + +E E    WQ+++++   L E+E Q K R+    +   ++R+ + +
Sbjct:   111 QNRRQEDQRRFELRDRQFEDEPERRRWQKQEQERELAEEEEQRKKRERFEQHYSRQYRDK 170

Query:   140 WVQLEKEQ-ESPRREPEQL----GEDVERRIFTPTSRWRDLEKAELSLVPAPSRTQSAHQ 194
              +L++++ E   R E EQL      G D E     F     + R   E+ EL              Q    +
Sbjct:   171 EQRLQRQELEERRAEEEQLRRRKGRDAEE--FIEEEQLRRREQQELKR-ELREEEQQRRE 227

Query:   195 SRRPHLPMSPSTQQPALGKQR 215
              R  H       ++   L ++R
Sbjct:   228 RREQHERALQEEEEQLLRQRR 248

 Score = 50 (7.5 bits), Expect = 2.2e+00, P = 8.9e-01
 Identities = 34/160 (21%), Positives = 67/160 (41%)

Query:   325 RLQSLRQEAINHVQIMKETEASYKAQNLYIFLENIDRL-QSLRLQAWTDKQKGLEEKHRE 383
             R +  R+E      Q+ +E E    + +         LE   +R  Q LR +     ++++ E++ R
Sbjct:   245 RQRRWREEPREQQQLRRELEEIREREQR---LEQEERREQQLRREQRLEQEERREQQLRR 301

Query:   384 CLSSMVTMFPKLQLEWNVHLNIP-EVTSPKPKKCKLPAASPRHIRPSGPTYKQPFLSRHR 442
             L   +     +L+ E        +   E    + K +L         R  R        ++   L+
Sbjct:   302 ELEEIREREQRLEQEERREQRLEQEERREQQLKRELEEIREREQRLEQEERREQLLAEEV 361

Query:   443 ACVPLQMARQQGKQMEAVWKTEVASSSYAIEKKTPASLPRDQ 484
                   + AR++G+ +   W+ ++ S + A + K   S PR Q
Sbjct:   362 R----EQARERGESLTRRWQRQLESEAGARQSKV-YSRPRRQ 398

 Score = 40 (6.0 bits), Expect = 1.9e-01, P = 1.7e-01
 Identities = 32/115 (27%), Positives = 47/115 (40%)

Query:   276 RKNLQLLSEESELRLPHYLRSKAL--ELTTTTMELGALRLQYLCHKYIFYRRL-QSLRQE 332
             R+  QLL E  E RL      R++ L  E        E    LR Q      K+      +L Q  +E
Sbjct:   959 REEEQLLQEREEERLRRQERARKLREEEQLLRREEQELR-QERDRKFREEEQLLQEREEE 1017

Query:   333 AINHVQI---MKETEASYKAQNLYI-FLENIDRLQSLRLQAWTDKQ-KGLEEKHRE 383
               +    +      +E E      + Q L      F +   DR     L  Q     +K+ K L    + R+
Sbjct:  1018 RLRRQERDRKFREEERQLRRQELEEQFRQERDRKFRLEEQIRQEKEEKQLRRQERD 1073

 Score = 37 (5.6 bits), Expect = 1.6e+00, P = 7.9e-01
 Identities = 27/108 (25%), Positives = 43/108 (39%)

Query:   276 RKNLQLLSEESELRLPHYLRSKAL---ELTTTTMELGALRLQYLCHKYIFYRRLQSLRQE 332
             R+  QLL E  E RL      R + L    E       E    LR Q      K      R   + L QE
Sbjct:   775 REEEQLLQEREEERLRRQERERKLREEEQLLQEREEERLRRQERERKL---REEEQLLQE 831

Query:   333 AINHVQIMKETEASYKAQNLYIFLENIDRLQSLRLQAWTDKQKGLEEKHRE 383
                        +E E      +      +     E      L++    R  +     ++++ L  ++  +E
Sbjct:   832 REEERLRRQERERKLREEEQLLRQEE-QELRQERARKLREEEQLLRQEEQE 881
```

Pedant information for DKFZphtes3_50n23, frame 1

Report for DKFZphtes3_50n23.1

```
[LENGTH]      499
[MW]          58885.69
[pI]          9.67
[KW]          All_Alpha
[KW]          LOW_COMPLEXITY     10.42 %

SEQ   MTVRSRVADVFGSKDTESLEPVLLPLVDRRFPKKWERPVAESLGHKDKDQEDYFQKGGLQ
SEG   ............................................................
PRD   ccccccceeeccccccccccceecccccccccccccchhhhhhcccccccccccccccce

SEQ   IKFHCSKQLSLESSRQVTSESQEEPWEEEFGREMRRQLWLEEEEMWQQRQKKWALLEQEH
SEG   ..................xxxxxxxxxx..xxxxxxxxxxxxxxxxxxx...........
PRD   eeeecchhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   QEKLRQWNLEDLAREQQRRWVQLEKEQESPRREPEQLGEDVERRIFTPTSRWRDLEKAEL
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccceeeccccccchhhhhhh

SEQ   SLVPAPSRTQSAHQSRRPHLPMSPSTQQPALGKQRPMSSVEFTYRPRTRRVPTKPKKSAS
SEG   ..........................................xxxxxxxxxxxxxxx...
PRD   hcccccccchhhhcccccccccccccccccccccccccceeeeeeccccccccccccceee

SEQ   FPVTGTSIRRLTWPSLQISPANIKKKVYHMDMEAQRKNLQLLSEESELRLPHYLRSKALE
SEG   ........................................xxxxxxxx............
PRD   eccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   LTTTTMELGALRLQYLCHKYIFYRRLQSLRQEAINHVQIMKETEASYKAQNLYIFLENID
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   RLQSLRLQAWTDKQKGLEEKHRECLSSMVTMFPKLQLEWNVHLNIPEVTSPKPKKCKLPA
SEG   ............................................................
PRD   hhhhhhhhhhhcchhhhhhhhhhhhhhhhccccchhhhhccccccccccccccccccccc

SEQ   ASPRHIRPSGPTYKQPFLSRHRACVPLQMARQQGKQMEAVWKTEVASSSYAIEKKTPASL
SEG   ............................................................
PRD   ccccccccccccccchhhhhhcchhhhhhhhhcchhhhhhhhhhhhhhhhhhhhcccccc

SEQ   PRDQLRGHPDIPRLLTLDV
SEG   ...................
PRD   ccccccccccccccccccc
```

(No Prosite data available for DKFZphtes3_50n23.1)

(No Pfam data available for DKFZphtes3_50n23.1)

DKFZphtes3_6b21
----

group: testes derived

DKFZphtes3_6b21 encodes a novel 781 amino acid protein without similarity to human KIAA0256 gene product.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to KIAA0256

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="356.3 cR from top of Chr9 linkage group"

Insert length: 3360 bp
Poly A stretch at pos. 3314, polyadenylation signal at pos. 3300

```
   1 GGCAAGCCGA CGGCCCGCTG CTGGCCTCCG TGACGCGGCC TCCTCCGCGC
  51 CTCGCGGCAT GGCGTCGGAG GGGCCGCGGG AGCCCGAAAG CGAGGGCATC
 101 AAGTTATCAG CAGATGTCAA ACCATTTGTC CCCAGATTTG CCGGGCTCAA
 151 TGTGGCATGG TTAGAGTCCT CAGAAGCATG TGTCTTCCCC AGCTCTGCAG
 201 CCACATACTA TCCGTTTGTT CAGGAACCAC CAGTGACAGA AATGTTTACT
 251 CAGTGCCTGG CTCCCAGTAT CTTTATAACC AACCCAGTTG TTACCGAGGT
 301 TTTCAAACAG TGAAGCATCG AAATGAGAAC ACATGCCCTC TCCCACAAGA
 351 AATGAAAGCT CTGTTTAAGA AGAAAACCTA TGATGAGAAA AAAACGTATG
 401 ATCAGCAAAA GTTTGACAGT GAAAGGGCTG ATGGAACTAT ATCATCTGAG
 451 ATAAAATCAG CTAGAGGTTC ACATCATTTG TCCATTTACG CTGAGAATAG
 501 TTTGAAATCA GATGGTTACC ATAAGCGAAC AGACAGGAAA TCCAGAATCA
 551 TTGCAAAAAA TGTATCTACC TCCAAACCTG AGTTTGAATT TACCACACTG
 601 GACTTTCCTG AACTGCAAGG TGCAGAGAAC AATATGTCAG AGATACAGAA
 651 GCAACCCAAG TGGGGACCTG TCCACTCTGT CTCTACCGAC ATTTCTCTTC
 701 TAAGAGAAGT AGTAAAACCA GCTGCAGTGT TATCAAAGGG TGAAATAGTG
 751 GTGAAAAATA ACCCAAATGA ATCTGTAACT GCTAATGCCG CTACCAATTC
 801 TCCTTCATGT ACAAGAGAGT TATCTTGGAC ACCAATGGGT TATGTTGTTC
 851 GACAGACATT ATCTACAGAA CTGTCAGCAG CCCCTAAAAA TGTTACTTCT
 901 ATGATAAACT TAAAGACCAT TGCTTCATCA GCAGATCCTA AAAATGTTAG
 951 TATACCATCT TCTGAAGCTT TATCTTCGGA TCCTTCCTAC AACAAAGAAA
1001 AACACATTAT TCATCCTACC CAAAAGTCTA AAGCATCACA AGGTAGTGAC
1051 CTTGAACAAA ATGAAGCCTC AAGAAAGAAT AAGAAAAAGA AAGAAAAATC
1101 TACATCAAAA TATGAAGTCC TGACAGTTCA AGAGCCTCCA AGGATTGAAG
1151 ATGCCGAGGA ATTTCCCAAC CTGGCAGTTG CATCTGAAAG AAGAGACAGA
1201 ATAGAGACAC CGAAATTTCA ATCTAAGCAG CAGCCACAGG ATAATTTTAA
1251 AAATAATGTA AAGAAGAGCC AGCTTCCAGT GCAGTTGGAC TTGGGGGGCA
1301 TGCTGACAGC CCTGGAGAAG AAGCAGCACT CTCAGCATGC AAAGCAGTCC
1351 TCCAAACCAG TGGTAGTCTC AGTTGGAGCA GTGCCAGTCC TTTCCAAAGA
1401 ATGTGCATCA GGGGAGAGAG GCCGCCGCAT GAGTCAAATG AAGACCCCGC
1451 ACAATCCCTT GGACTCCAGC GCCCCACTGA TGAAGAAAGG GAAGCAGAGG
1501 GAGATCCCCA AGGCCAAGAA GCCAACCTCA CTGAAGAAGA TTATTTTGAA
1551 AGAACGGCAA GAGAGAAAGC AGCGTCTCCA AGAAAATGCT GTGAGTCCAG
1601 CTTTTACCAG TGATGACACA CAAGATGGAG AGAGTGGTGG TGATGACCAG
1651 TTTCCCGAGC AGGCAGAGCT GTCAGGGCCA GAGGGGATGG ACGAACTGAT
1701 CTCCACTCCT TCGGTTGAGG ACAAGTCTGA AGAGCCACCA GGCACAGAGC
1751 TCCAGAGGGA CACAGAGGCC TCCCACCTTG CTCCCAATCA CACCACCTTC
1801 CCTAAGATCC ACAGCCGCAG ATTCAGGGAT TACTGCAGCC AGATGCTTAG
1851 TAAAGAAGTG GATGCTTGTG TTACCGACCT ACTCAAAGAA CTGGTCCGTT
1901 TCCAAGACCG TATGTACCAG AAAGATCCAG TCAAGGCCAA GACTAAACGT
1951 CGACTTGTGT TGGGGTTGAG GGAGGTTCTC AAACACCTGA AGCTCAAAAA
2001 ACTGAAATGT GTCATTATTT CTCCCAACTG TGAGAAGATA CAGTCAAAAG
2051 GTGGGCTGGA TGACACTTTG CACACAATTA TTGATTATGC CTGTGAGCAG
2101 AACATTCCCT TTGTGTTTGC TCTCAACCGC AAAGCTCTGG GGCGCAGTTT
2151 GAATAAGGCA GTTCCTGTCA GTGTGGTGGG GATCTTCAGC TATGATGGGG
2201 CCCAGGATCA GTTCCACAAG ATGGTTGAGC TGACAGTGGC GGCCCGACAG
2251 GCGTACAAGA CCATGCTGGA GAATGTGCAG CAGGAGCTGG TGGGAGAGCC
2301 CAGGCCTCAG GCACCTCCCA GCCTACCCAC ACAGGGCCCC AGCTGCCCTG
2351 CAGAAGATGG CCCCCCAGCC CTGAAAGAAA AAGAAGAGCC ACACTACATT
2401 GAAATCTGGA AAAAACATCT GGAAGCATAC AGTGGATGTA CCCTGGAGCT
2451 AGAAGAATCC TTGGAGGCTT CAACCTCTCA AATGATGAAT TTGAATTTAT
2501 GAGAGTTCTT GCCTGTGTGT CTGTATTTTG GGTAAGGAGG GGAGGTCTGA
2551 AAAAGACTTT GGGGCTTTTT CTTCTGTTTT TCATGACAAT GTAATTTGTG
2601 TAACTGTTGA ATCTGGAAAT TGATCAGCAT TAAAGGGCAC ATGAAGCAGT
2651 GTCTGCAGGC GTTCAGTGCT GCGGAGCCTG TTAAAGGTCA CTCAGATGTG
```

```
2701 CAGGTGTTAA TCTTCTCTAA AAGCCTGGTT ATACAGCTCT GGCTTTCTGA
2751 GCACACTACG GATCTGGAAA ATACTGGAAA ATGTGATACT TAGAATACTT
2801 TGGCTGCTAA GGAAACTTCC TCTCCATTGC AGAATAGCTG AGCCAAGTGA
2851 GTGAGTTTGC AGAAAGCAGG TGGTGAGCTC CTGCCTGCTG GAGGTTGCCA
2901 TGGAGGGCCA TTCCTGCCCG GCAACAGCAC CGTCCTGCAG GGAGCCACTT
2951 GGCAGAAGGG TGCAGGGCTG CTGGTGTCAG AGCAAGAGGG CTACAGGGAA
3001 AGGGCCCTTT CTCAGGGGAT GTAGCTTTTT TAAAAGATTT GGGAACACTT
3051 GGAGGATTTG CTAAAATGAG CCTCAGAAGG AAAATTGGTT TTCTAACCTG
3101 TGACTTTTTG AAATGAATTA TTCCTTTCAG TCTTTATTTT TCAAAGAAAC
3151 AATGTGTATT GAAGTACCTA GATTTGTTTG ATAATCAACA AATCTTTCCT
3201 TTTTCAATGA ACATATTCTG AATGTGGTTT CTGTCTTAGA CCAGGAGGAC
3251 AGAGTTTGCT TTCATATTTT CCCTGTAAGT AAGAGGGCTT ATTTATTTTA
3301 AATAAAGAGT AATTATTAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAAAAAAAAA
```

BLAST Results

Entry HS773347 from database EMBL:
human STS WI-18160.
Score = 813, P = 2.9e-30, identities = 167/171

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 157 bp to 2499 bp; peptide length: 781
Category: similarity to known protein

```
  1 MVRVLRSMCL PQLCSHILSV CSGTTSDRNV YSVPGSQYLY NQPSCYRGFQ
 51 TVKHRNENTC PLPQEMKALF KKKTYDEKKT YDQQKFDSER ADGTISSEIK
101 SARGSHHLSI YAENSLKSDG YHKRTDRKSR IIAKNVSTSK PEFEFTTLDF
151 PELQGAENNM SEIQKQPKWG PVHSVSTDIS LLREVVKPAA VLSKGEIVVK
201 NNPNESVTAN AATNSPSCTR ELSWTPMGYV VRQTLSTELS AAPKNVTSMI
251 NLKTIASSAD PKNVSIPSSE ALSSDPSYNK EKHIIHPTQK SKASQGSDLE
301 QNEASRKNKK KKEKSTSKYE VLTVQEPPRI EDAEEFPNLA VASERRDRIE
351 TPKFQSKQQP QDNFKNNVKK SQLPVQLDLG GMLTALEKKQ HSQHAKQSSK
401 PVVVSVGAVP VLSKECASGE RGRRMSQMKT PHNPLDSSAP LMKKGKQREI
451 PKAKKPTSLK KIILKERQER KQRLQENAVS PAFTSDDTQD GESGGDDQFP
501 EQAELSGPEG MDELISTPSV EDKSEEPPGT ELQRDTEASH LAPNHTTFPK
551 IHSRRFRDYC SQMLSKEVDA CVTDLLKELV RFQDRMYQKD PVKAKTKRRL
601 VLGLREVLKH LKLKKLKCVI ISPNCEKIQS KGGLDDTLHT IIDYACEQNI
651 PFVFALNRKA LGRSLNKAVP VSVVGIFSYD GAQDQFHKMV ELTVAARQAY
701 KTMLENVQQE LVGEPRPQAP PSLPTQGPSC PAEDGPPALK EKEEPHYIEI
751 WKKHLEAYSG CTLELEESLE ASTSQMMNLN L
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_6b21, frame 1

SWISSPROT:Y256_HUMAN HYPOTHETICAL PROTEIN KIAA0256., N = 1, Score = 786, P = 3.6e-78

TREMBL:PFMAL3P3_15 gene: "MAL3P3.15"; Plasmodium falciparum MAL3P3, N = 2, Score = 161, P = 5.1e-10

TREMBL:RNNFLH_1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N = 1, Score = 150, P = 9.1e-07

>SWISSPROT:Y256_HUMAN HYPOTHETICAL PROTEIN KIAA0256.
Length = 635

HSPs:

Score = 786 (117.9 bits), Expect = 3.6e-78, P = 3.6e-78
Identities = 190/424 (44%), Positives = 263/424 (62%)

```
Query:   369 KKSQLPVQLDLGGMLTALEKKQHSQHAKQ--SSKPVVVSVGAVPVLSKECASGERGRRMS 426
             KK++ PVQLDLG ML ALEK+Q + A+Q +++P+ +V         + ++ +    S
Sbjct:    16 KKNKTPVQLDLGDMLAALEKQQQAMKARQITNTRPLSYTVVTAASFHTKDSTNRKPLTKS 75

Query:   427 Q-MKTPHNPLDSSAPLMKKGKQREIPKAKKPTSLKKIILKERQERKQRLQENAVSPAFTS 485
             Q  T  N +D ++    KKGK++EI K K+PT+LKK+ILKER+E+K RL   +      S
Sbjct:    76 QPCLTSFNSVDIASSKAKKGKEKEIAKLKRPTALKKVILKEREEKKGRLTVD--HNLLGS 133

Query:   486 DDTQDGESGGDDQFPEQAELSGPEGMDELISTPSVEDKSEEPPG--TELQRDTEASHL-- 541
             ++  +       D  P++        G+ + S  S+   S+  P   T + + + AS
Sbjct:   134 EEPTEMHLDFIDDLPQEIVSQEDTGLS-MPSDTSLSPASQNSPYCMTPVSQGSPASSGIG 192

Query:   542 APN-HTTFPKIHSRRFRDYCSQMLSKEVDACVTDLLKELVRFQDRMYQKDPVKAKTKRRL 600
             +P    +T  KIHS+RFR+YC+Q+L KE+D CVT LL+ELV FQ+R+YQKDPV+AK +RRL
Sbjct:   193 SPMASSTITKIHSKRFREYCNQVLCKEIDECVTLLLQELVSFQERIYQKDPVRAKARRRL 252

Query:   601 VLGLREVLKHLKLKKLKCVIISPNCEKIQSKGGLDDTLHTIIDYACEQNIPFVFALNRKA 660
             V+GLREV KH+KL K+KCVIISPNCEKIQSKGGLD+ L+ +I  A EQ IPFVFAL RKA
Sbjct:   253 VMGLREVTKHMKLNKIKCVIISPNCEKIQSKGGLDEALYNVIAMAREQEIPFVFALGRKA 312

Query:   661 LGRSLNKAVPVSVVGIFSYDGAQDQFHKMVELTVAARQAYKTMLENVQQELVGEPRP--- 717
             LGR +NK VPVSVVGIF+Y GA+  F+K+VELT  AR+AYK M+  ++QE   E
Sbjct:   313 LGRCVNKLVPVSVVGIFNYFGAESLFNKLVELTEEARKAYKDMVAAMEQEQAEEALKNVK 372

Query:   718 QAPPSLP-TQGPS-------CPAEDGPPALKEKEEPHYIEIWKKHLEAYSGCTL---ELE 766
             + P  +  ++ PS       C    P + E  E  Y  W+  +E   G       E E
Sbjct:   373 KVPHHMGHSRNPSAASAISFCSVISEP--ISEVNEKEYETNWRNMVETSDGLEASENEKE 430

Query:   767 ESLEASTSQ 775
              S + STS+
Sbjct:   431 VSCKHSTSE 439
```

Pedant information for DKFZphtes3_6b21, frame 1

Report for DKFZphtes3_6b21.1

```
[LENGTH]    781
[MW]        87393.44
[pI]        8.94
[HOMOL]     SWISSPROT:Y256_HUMAN HYPOTHETICAL PROTEIN KIAA0256. 4e-75
[PROSITE]   MYRISTYL        4
[PROSITE]   AMIDATION       1
[PROSITE]   CAMP_PHOSPHO_SITE      3
[PROSITE]   CK2_PHOSPHO_SITE       16
[PROSITE]   TYR_PHOSPHO_SITE       4
[PROSITE]   PKC_PHOSPHO_SITE       16
[PROSITE]   ASN_GLYCOSYLATION      6
[KW]        Alpha_Beta
[KW]        LOW_COMPLEXITY      8.45 %

SEQ    MVRVLRSMCLPQLCSHILSVCSGTTSDRNVYSVPGSQYLYNQPSCYRGFQTVKHRNENTC
SEG    ............................................................
PRD    ccceeeeecccceeeeeeeeeeeccccccccccccccccccccccceeeceeeeeeeccccc

SEQ    PLPQEMKALFKKKTYDEKKTYDQQKFDSERADGTISSEIKSARGSHHLSIYAENSLKSDG
SEG    .........xxxxxxxxxxx........................................
PRD    ccchhhhhhhhhhhcchhhhhhhhhhhcccccchhhhhhhcccceeeeeeeeeecccccc

SEQ    YHKRTDRKSRIIAKNVSTSKPEFEFTTLDFPELQGAENNMSEIQKQPKWGPVHSVSTDIS
SEG    ............................................................
PRD    ccccchhhhheeeccccccccceeeccccccccccchhhhhhccccccccccceeecchh

SEQ    LLREVVKPAAVLSKGEIVVKNNPNESVTANAATNSPSCTRELSWTPMGYVVRQTLSTELS
SEG    ............................................................
PRD    hhhhhhheeeeeccccceeeeccccceeeeeeccccccccceeeeecccceeeeeecccccc

SEQ    AAPKNVTSMINLKTIASSADPKNVSIPSSEALSSDPSYNKEKHIIHPTQKSKASQGSDLE
SEG    ............................................................
PRD    cccccceeeehhhhhhcccccceeeecccccccccccccccceeechhhhhhhcccccch

SEQ    QNEASRKNKKKKEKSTSKYEVLTVQEPPRIEDAEEFPNLAVASERRDRIETPKFQSKQQP
SEG    ....xxxxxxxxxxxxxx..........................................
PRD    hhhhccccccccccccceeeeeecccccchhhhhhcchhhhhhhhhhhhcccccccccc

SEQ    QDNFKNNVKKSQLPVQLDLGGMLTALEKKQHSQHAKQSSKPVVVSVGAVPVLSKECASGE
SEG    ..................................xxxxxxxxxxxxxxxx......
PRD    cccccccccccccccccchhhhhhhhhhhhhhhhhhhcccceeeeeeeeeeeeeccccc
```

```
SEQ   RGRRMSQMKTPHNPLDSSAPLMKKGKQREIPKAKKPTSLKKIILKERQERKQRLQENAVS
SEG   ............................................................
PRD   chhhhhcccccccccccccchhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhcc

SEQ   PAFTSDDTQDGESGGDDQFPEQAELSGPEGMDELISTPSVEDKSEEPPGTELQRDTEASH
SEG   ............................................................
PRD   cccccccccccccccccchhhhhccccccceeeeccccccccccccccccccccccccccc

SEQ   LAPNHTTFPKIHSRRFRDYCSQMLSKEVDACVTDLLKELVRFQDRMYQKDPVKAKTKRRL
SEG   ............................................................
PRD   ccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhhhh

SEQ   VLGLREVLKHLKLKKLKCVIISPNCEKIQSKGGLDDTLHTIIDYACEQNIPFVFALNRKA
SEG   .......xxxxxxxxxx...........................................
PRD   hhhhhhhhhhhhhhhhheeeeccccccccccccccchhhhhhhhhhhhcccceeeecccc

SEQ   LGRSLNKAVPVSVVGIFSYDGAQDQFHKMVELTVAARQAYKTMLENVQQELVGEPRPQAP
SEG   ............................................................
PRD   cccccccceeeeeeeeeecccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccc

SEQ   PSLPTQGPSCPAEDGPPALKEKEEPHYIEIWKKHLEAYSGCTLELEESLEASTSQMMNLN
SEG   .........................................xxxxxxxxxxxxx......
PRD   cccccccccccccccchhhhhhcccceeeehhhhhhhhchhhhhhhhhhhhhhhhccccc

SEQ   L
SEG   .
PRD   c
```

Prosite for DKFZphtes3_6b21.1

PS00001	135->139	ASN_GLYCOSYLATION	PDOC00001
PS00001	159->163	ASN_GLYCOSYLATION	PDOC00001
PS00001	204->208	ASN_GLYCOSYLATION	PDOC00001
PS00001	245->249	ASN_GLYCOSYLATION	PDOC00001
PS00001	263->267	ASN_GLYCOSYLATION	PDOC00001
PS00001	544->548	ASN_GLYCOSYLATION	PDOC00001
PS00004	71->75	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	423->427	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	454->458	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	51->54	PKC_PHOSPHO_SITE	PDOC00005
PS00005	88->91	PKC_PHOSPHO_SITE	PDOC00005
PS00005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	125->128	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	288->291	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00005	316->319	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	398->401	PKC_PHOSPHO_SITE	PDOC00005
PS00005	458->461	PKC_PHOSPHO_SITE	PDOC00005
PS00005	553->556	PKC_PHOSPHO_SITE	PDOC00005
PS00005	596->599	PKC_PHOSPHO_SITE	PDOC00005
PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00006	74->78	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	297->301	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	384->388	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	493->497	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	640->644	CK2_PHOSPHO_SITE	PDOC00006
PS00006	702->706	CK2_PHOSPHO_SITE	PDOC00006
PS00007	581->588	TYR_PHOSPHO_SITE	PDOC00007
PS00007	740->748	TYR_PHOSPHO_SITE	PDOC00007
PS00007	740->748	TYR_PHOSPHO_SITE	PDOC00007
PS00007	73->82	TYR_PHOSPHO_SITE	PDOC00007
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00008	380->386	MYRISTYL	PDOC00008

PS00008	633->639	MYRISTYL	PDOC00008
PS00009	421->425	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_6b21.1)

DKFZphtes3_6c11
----

group: signal transduction

DKFZphtes3_6c11 encodes a novel 1025 amino acid protein with similarity to A. ambisexualis antheridiol steroid receptor.

The novel protein is a putative steroid receptor. It shares similarity with yeast YNL132w and contains the ATP/GTP-binding site motif A (P-loop) and RGD site, similar to the A. ambisexualis antheridiol steroid receptor.

The new protein can find application in modulating/blocking the expression of genes controlled by this receptor.

strong similarity to YNL132w

strong similarity to S.pombe/YDK9_SCHPO, S.cerevisiae/YNL132w, C.elegans/F55A12.8

Sequenced by BMFZ

Locus: unknown

Insert length: 3966 bp
Poly A stretch at pos. 3890, polyadenylation signal at pos. 3873

```
   1 GCTGTGCCTT CTCTTTCGGA GTTGTTCCGT GCTCCCACGT GCTTCCCCTT
  51 CTCCACTGGC TGGGATCCCC CGGGCTCGGG GCGCAGTAAT AATTTTTCAC
 101 CATGCATCGG AAAAAGGTGG ATAACCGAAT CCGGATTCTC ATTGAGAATG
 151 GAGTAGCTGA GCGGCAAAGA TCTCTCTTTG TTGTAGTTGG GGATCGAGGA
 201 AAAGATCAGG TGGTAATACT TCATCACATG TTATCCAAAG CAACTGTGAA
 251 GGCTCGGCCT TCAGTGCTGT GGTGTTATAA GAAAGAGCTG GGGTTTAGCA
 301 GTCACCGGAA GAAAAGAATG CGACAGCTGC AGAAGAAAAT AAAGAATGGA
 351 ACACTGAACA TAAAGCAGGA CGACCCCTTT GAACTCTTCA TAGCAGCCAC
 401 AAACATTCGC TACTGCTACT ACAACGAGAC CCACAAGATC CTGGGCAATA
 451 CCTTCGGCAT GTGTGTGCTG CAGGATTTTG AAGCCTTAAC TCCAAACTTG
 501 CTGGCCAGGA CTGTAGAAAC AGTGGAAGGT GGTGGGCTAG TGGTCATCCT
 551 CCTACGGACC ATGAACTCAC TCAAGCAATT GTACACAGTG ACTATGGATG
 601 TGCATTCCAG GTACAGAACT GAGGCCCATC AGGATGTGGT GGGAAGATTT
 651 AATGAAAGGT TTATTCTGTC TCTGGCCTCT TGTAAGAAGT GTCTCGTCAT
 701 TGATGACCAG CTCAACATCC TGCCCATCTC CTCCCACGTT GCCACCATGG
 751 AGGCCCTGCC TCCCCAGACT CCGGATGAGA GTCTTGGTCC TTCTGATCTG
 801 GAGCTGAGGG AGTTGAAGGA GAGCTTGCAG GACACCCAGC CTGTGGGTGT
 851 GTTGGTGGAC TGCTGTAAGA CTCTAGACCA GGCCAAAGCT GTCTTGAAAT
 901 TTATCGAGGG CATCTCTGAA AAGACCCTGA GGAGTACTGT TGCACTCACA
 951 GCTGCTCGAG GACGGGGAAA ATCTGCAGCC CTGGGATTGG CGATTGCTGG
1001 GGCGGTGGCA TTTGGGTACT CCAATATCTT TGTTACCTCC CCAAGCCCTG
1051 ATAACCTCCA TACTCTGTTT GAATTTGTAT TTAAAGGATT TGATGCTCTG
1101 CAATATCAGG AACATCTGGA TTATGAGATT ATCCAGTCTC TAAATCCTGA
1151 ATTTAACAAA GCAGTGATCA GAGTGAATGT ATTTCGAGAA CACAGGCAGA
1201 CTATTCAGTA TATACATCCT GCAGATGCTG TGAAGCTGGG CCAGGCTGAA
1251 CTAGTTGTGA TTGATGAAGC TGCCGCCATC CCCCTCCCCT TGGTGAAGAG
1301 CCTACTTGGC CCCTACCTTG TTTTCATGGC ATCCACCATC AATGGCTATG
1351 AGGGCACTGG CCGGTCACTG TCCCTCAAGC TAATTCAGCA GCTCCGTCAA
1401 CAGAGCGCCC AGAGCCAGGT CAGCACCACT GCTGAGAATA AGACCACGAC
1451 GACAGCCAGA TTGGCATCAG CGCGGACACT GCATGAGGTT TCCCTCCAGG
1501 AGTCAATCCG ATACGCCCCT GGGGATGCAG TGGAGAAGTG GCTGAATGAC
1551 TTGCTGTGCC TGGATTGCCT CAACATCACT CGGATAGTCT CAGGCTGCCC
1601 CTTGCCTGAA GCTTGTGAAC TGTACTATGT TAATAGAGAT ACCCTCTTTT
1651 GCTACCACAA GGCCTCTGAA GTTTTCCTCC AACGGCTTAT GGCCCTCTAC
1701 GTGGCTTCTC ACTACAAGAA CTCTCCCAAT GATCTCCAGA TGCTCTCCGA
1751 TGCACCTGCT CACCATCTCT TCTGCCTTCT GCCTCCTGTG CCCCCCACCC
1801 AGAATGCCCT TCCAGAAGTG CTTGCTGTTA TCCAGGTGTG CCTTGAAGGG
1851 GAGATTTCTC GCCAGTCCAT CTTGAACAGT CTGTCTCGAG GCAAGAAGGC
1901 TTCAGGGGAC CTGATTCCAT GGACAGTGTC AGAACAGTTC CAAGATCCAG
1951 ACTTTGGTGG TCTGTCTGGT GGAAGGGTCG TTCGCATTGC TGTTCACCCA
2001 GATTATCAAG GGATGGGCTA TGGCAGCCGT GCTCTGCAGC TGCTGCAGAT
2051 GTACTATGAA GGCAGGTTTC CTTGTCTGGA GGAAAAGGTC CTTGAGACAC
2101 CACAGGAAAT TCACACCGTA AGCAGCGAGG CTGTCAGCTT GTTGGAAGAG
2151 GTCATCACTC CCCGGAAGGA CCTGCCTCCT TTACTCCTCA AATTGAATGA
2201 GAGGCCTGCC GAACGCCTGG ATTACCTGGG TGTTTCCTAT GGCTTGACCC
2251 CCAGGCTCCT CAAGTTCTGG AAACGAGCTG GATTTGTTCC TGTTTATCTG
2301 AGACAGACCC CGAATGACCT GACCGGAGAG CACTCGTGCA TCATGCTGAA
2351 GACGCTCACT GATGAGGATG AGGCTGACCA GGGAGGCTGG CTTGCAGCCT
2401 TCTGGAAAGA TTTCCGACGG CGGTTCCTAG CCTTGCTCTC CTACCAGTTC
2451 AGTACCTTCT CTCCTTCCCT GGCTCTGAAC ATCATTCAGA ACAGGAACAT
2501 GGGGAAGCCA GCCCAGCCTG CCCTGAGCCG GGAGGAGCTG GAAGCACTCT
```

```
2551 TCCTCCCCTA TGACCTGAAG CGGCTGGAGA TGTATTCACG GAATATGGTG
2601 GACTATCACC TCATCATGGA CATGATCCCG GCCATCTCTC GCATCTATTT
2651 CCTGAACCAG CTGGGGGACC TGGCCCTGTC TGCGGCTCAG TCGGCTCTTC
2701 TCTTGGGGAT TGGCCTGCAG CATAAGTCTG TGGACCAGCT GGAAAAGGAG
2751 ATTGAGCTGC CCTCGGGCCA GTTGATGGGA CTTTTCAACC GGATCATCCG
2801 CAAAGTTGTG AAGCTATTTA ATGAAGTTCA GGAAAAGGCC ATTGAGGAGC
2851 AGATGGTGGC AGCGAAGGAT GTGGTCATGG AGCCCACGAT GAAGACCCTC
2901 AGTGACGACC TAGATGAAGC AGCAAAGGAA TTTCAGGAGA AACACAAGAA
2951 GGAAGTAGGG AAGCTGAAGA GCATGGACCT CTCTGAATAC ATAATCCGTG
3001 GGGACGATGA AGAGTGGAAT GAAGTTTTGA ACAAAGCTGG GCCGAACGCC
3051 TCGATCATCA GCCTGAAAAG TGACAAGAAA AGGAAGTTAG AGGCCAAACA
3101 AGAACCCAAA CAGAGCAAGA AGTTGAAGAA CAGAGAGACA AAGAACAAAA
3151 AAGATATGAA ACTGAAGCGG AAGAAATAGT GAAGAGAAAC TCGGGCATCT
3201 GTGTTTGATC ATGGGAAGAT ACTCTCACTA ACTGAACCCT CTCTGGCTGG
3251 ACTGTTAAAA GCAACGAGAG GCCCCGGCAC ACCTGGAAGC TGGCCGCGAA
3301 TTCGGCCTCT GGGCCTGTGT GTCTGTGAGC TCAACCTGGC TAAAGGCAGA
3351 GTCACTCCCA AATGGGTCTC TTTAGAACTT GATGGCTGGG CACTGCCATC
3401 TCTAGAATTG CCACGAGTCT CTCTCTTCCT GCCCAGTCCA GGGCCCTCCT
3451 TTCCTATAAG TTCATATTTT GCTTTGAGCC AGCTTTTTAG TCTCATTCCC
3501 ACACATGTGG AAGCCACGTT GCCTCTCGAC CGCCTGAGGC CCTTAAGTAC
3551 ATCGCTTTCT GGTGGTGCCC AGGAGGCTGC TGCTGGGCCG CTGGGTCTCT
3601 CTTTGTGGAC TTGTACCTGG AGCAGGAGGA ACTCCAGTCC GTCCCGGCAT
3651 CCATGGCAGC CCGCGGTTAG GTGCGCCAGG GTTTGCTGAT GTTGTCTTGT
3701 GCTGTTCCAC TCTTGGCTCC AGCAGACCCA CTGTCCCAGA AAAGCCTGAT
3751 CCTGTAGTTT ATGTAGAATG CCACATCTGC GTCCTCAAGA CCTGTTTCAT
3801 CCATTTGGGA AAAGATGTTG GGAAAGGCCA CTTTGCTCGC AGGGGTGAGG
3851 GGAAGGATAG AGAATCTATT TTTAATAAAT AACATTCTAG AATGAAAAAA
3901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3951 AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 102 bp to 3176 bp; peptide length: 1025
Category: similarity to unknown protein
Classification: unclassified
Prosite motifs: RGD (966-969)
ATP_GTP_A (284-292)

```
   1 MHRKKVDNRI RILIENGVAE RQRSLFVVVG DRGKDQVVIL HHMLSKATVK
  51 ARPSVLWCYK KELGFSSHRK KRMRQLQKKI KNGTLNIKQD DPFELFIAAT
 101 NIRYCYYNET HKILGNTFGM CVLQDFEALT PNLLARTVET VEGGGLVVIL
 151 LRTMNSLKQL YTVTMDVHSR YRTEAHQDVV GRFNERFILS LASCKKCLVI
 201 DDQLNILPIS SHVATMEALP PQTPDESLGP SDLELRELKE SLQDTQPVGV
 251 LVDCCKTLDQ AKAVLKFIEG ISEKTLRSTV ALTAARGRGK SAALGLAIAG
 301 AVAFGYSNIF VTSPSPDNLH TLFEFVFKGF DALQYQEHLD YEIIQSLNPE
 351 FNKAVIRVNV FREHRQTIQY IHPADAVKLG QAELVVIDEA AIPLPLVKS
 401 LLGPYLVFMA STINGYEGTG RSLSLKLIQQ LRQQSAQSQV STTAENKTTT
 451 TARLASARTL HEVSLQESIR YAPGDAVEKW LNDLLCLDCL NITRIVSGCP
 501 LPEACELYYV NRDTLFCYHK ASEVFLQRLM ALYVASHYKN SPNDLQMLSD
 551 APAHHLFCLL PPVPPTQNAL PEVLAVIQVC LEGEISRQSI LNSLSRGKKA
 601 SGDLIPWTVS EQFQDPDFGG LSGGRVVRIA VHPDYQGMGY GSRALQLLQM
 651 YYEGRFPCLE EKVLETPQEI HTVSSEAVSL LEEVITPRKD LPPLLLKLNE
 701 RPAERLDYLG VSYGLTPRLL KFWKRAGFVP VYLRQTPNDL TGEHSCIMLK
 751 TLTDEDEADQ GGWLAAFWKD FRRRFLALLS YQFSTFSPSL ALNIIQNRNM
 801 GKPAQPALSR EELEALFLPY DLKRLEMYSR NMVDYHLIMD MIPAISRIYF
 851 LNQLGDLALS AAQSALLLGI GLQHKSVDQL EKEIELPSGQ LMGLFNRIIR
 901 KVVKLFNEVQ EKAIEEQMVA AKDVVMEPTM KTLSDDLDEA AKEFQEKHKK
 951 EVGKLKSMDL SEYIIRGDDE EWNEVLNKAG PNASIISLKS DKKRKLEAKQ
1001 EPKQSKKLKN RETKNKKDMK LKRKK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_6c11, frame 3

TREMBL:CEAF3130_4 gene: "F55A12.8"; Caenorhabditis elegans cosmid F55A12., N = 1, Score = 2782, P = 1.1e-289

PIR:S55151 probable membrane protein YNL132w - yeast (Saccharomyces cerevisiae), N = 2, Score = 2549, P = 3.5e-273

SWISSPROT:YXX1_ACHAM HYPOTHETICAL PROTEIN (FRAGMENT)., N = 1, Score = 1013, P = 3.2e-102

SWISSPROT:YDK9_SCHPO HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME I., N = 1, Score = 2843, P = 3.8e-296

```
>SWISSPROT:YDK9_SCHPO HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME
     I.
         Length = 1,033

 HSPs:

 Score = 2843 (426.6 bits), Expect = 3.8e-296, P = 3.8e-296
 Identities = 576/1033 (55%), Positives = 750/1033 (72%)

Query:     1 MHRKKVDNRIRILIENGVAERQRSLFVVVGDRGKDQVVILHHMLSKATVKARPSVLWCYK 60
             M +K +D+RI  LI+NG  E+QRS FVVVGDR +DQVV LH +LS++ V ARP+VLW YK
Sbjct:     1 MPKKALDSRIPTLIKNGCQEKQRSFFVVVGDRARDQVVNLHWLLSQSKVAARPNVLWMYK 60

Query:    61 KEL-GFSSHRKKRMRQLQKKIKNGTLNIKQDDPFELFIAATNIRYCYYNETHKILGNTFG 119
             K+L GF+SHRKKR  +++K+IK G  +   +DPFELF + TNIRYCYY E+ KILG T+G
Sbjct:    61 KDLLGFTSHRKKRENKIKKEIKRGIRDPNSEDPFELFCSITNIRYCYYKESEKILGQTYG 120

Query:   120 MCVLQDFEALTPNLLARTVETVEGGGLVVILLRTMNSLKQLYTVTMDVHSRYRTEAHQDV 179
             M VLQDFEALTPNLLART+ETVEGGG+VV+LL  +NSLKQLYT++MD+HSRYRTEAH DV
Sbjct:   121 MLVLQDFEALTPNLLARTIETVEGGGIVVLLLHKLNSLKQLYTMSMDIHSRYRTEAHSDV 180

Query:   180 VGRFNERFILSLASCKKCLVIDDQLNILPISSHVATMEALPPQTPDESLGPSDLELRELK 239
               RFNERFILSL +C+ CLVIDD+LN+LPIS    ++ALPP +++    +   ++EL+
Sbjct:   181 TARFNERFILSLGNCENCLVIDDELNVLPISGG-KNVKALPPTLEEDN--STQNSIKELQ 237

Query:   240 ESLQDTQPVGVLVDCCKTLDQAKAVLKFIEGISEKTLRSTVALTAARGRGKSAALGLAIA 299
             ESL +  P G LV  KTLDQA+AVL F+E I EK+L+ TV+LTA RGRGKSAALGLAIA
Sbjct:   238 ESLGEDHPAGALVGVTKTLDQARAVLTFVESIVEKSLKGTVSLTAGRGRGKSAALGLAIA 297

Query:   300 GAVAFGYSNIFVTSPSPDNLHTLFEFVFKGFDALQYQEHLDYEIIQSLNPEFNKAVIRVN 359
              A+A GYSNIF+TSPSP+NL TLFEF+FKGFDAL Y+EH+DY+IIQS NP ++ A++RVN
Sbjct:   298 AAIAHGYSNIFITSPSPENLKTLFEFIFKGFDALNYEEHVDYDIIQSTNPAYHNAIVRVN 357

Query:   360 VFREHRQTIQYIHPADAVKLGQAELVVIDEAAAIPLPLVKSLLGPYLVFMASTINGYEGT 419
             +FR+HRQTIQYI P D+  LGQAELVVIDEAAAIPLPLV+ L+GPYLVFMASTINGYEGT
Sbjct:   358 IFRDHRQTIQYISPEDSNVLGQAELVVIDEAAAIPLPLVRKLIGPYLVFMASTINGYEGT 417

Query:   420 GRSLSLKLIQQLRQQSAQSQVSTTAENKTTTTARLASARTLHEVSLQESIRYAPGDAVEK 479
             GRSLSLKL+QQLR+QS     S +  NK+ + + + S RTL E+SL E IRYA GD +E
Sbjct:   418 GRSLSLKLLQQLREQSRI--YSGSGNNKSDSQSHI-SGRTLKEISLDEPIRYAMGDRIEL 474

Query:   480 WLNDLLCLDCLN-ITRIVS-GCPLPEACELYYVNRDTLFCYHKASEVFLQRLMALYVASH 537
             WLN LLCLD  + ++R+ + G P P  C LY V+RDTLF YH  SE FLQR+M+LYVASH
Sbjct:   475 WLNKLLCLDAASYVSRMATQGFPHPSECSLYRVSRDTLFSYHPISEAFLQRMMSLYVASH 534

Query:   538 YKNSPNDLQMLSDAPAHHLFCLLPPVPPTQNALPEVLAVIQVCLEGEISRQSILNSLSRG 597
             YKNSPNDLQ++SDAPAH LF LLPPV      LP+ + VIQ+ LEG ISR+SI+NSLSRG
Sbjct:   535 YKNSPNDLQLMSDAPAHQLFVLLPPVDLKNPKLPDPICVIQLALEGSISRESIMNSLSRG 594

Query:   598 KKASGDLIPWTVSEQFQDPDFGGLSGGRVVRIAVHPDYQGMGYGSRALQLLQMYYEGRFP 657
             ++A GDLIPW +S+QFQD +F  L G R+VRIAV P++  MGYG+RA+QLL  Y+EG+F
Sbjct:   595 QRAGGDLIPWLISQQFQDENFAALGGARIVRIAVSPEHVKMGYGTRAMQLLHEYFEGKFI 654

Query:   658 CLEEKVLETPQEIHTVSSEAV---SLLEEVITPR--KDLPPLLLKLNERPAERLDYLGVS 712
                E+        +  + E +   +L  E I  R  K +PPLLLKL+E   E L Y+GVS
Sbjct:   655 SASEEFKAVKHSLKRIGDEEIENTALQTEKIHVRDAKTMPPLLLKLSELQPEPLHYVGVS 714

Query:   713 YGLTPRLLKFWKRAGFVPVYLRQTPNDLTGEHSCIMLKTLTDEDEADQGGWLAAFWKDFR 772
             YGLTP L KFWKR G+ P+YLRQT NDLTGEH+C+ML+ L  D     WL AF ++F
Sbjct:   715 YGLTPSLQKFWKREGYCPLYLRQTANDLTGEHTCVMLRVLEGRDSE----WLGAFAQNFY 770

Query:   773 RRFLALLSYQFSTFSPSLALNIIQNRNMGKP----AQPALSREELEALFLPYDLKRLEMY 828
             RRFL+LL YQF  F+  AL+++   N G       +  L+ EE+  +F  YDLKRLE Y
Sbjct:   771 RRFLSLLGYQFREFAAITALSVLDACNNGTKYVVNSTSKLTNEEINNVFESYDLKRLESY 830
```

```
Query:   829 SRNMVDYHLIMDMIPAISRIYFLNQLGD-LALSAAQSALLLGIGLQHKSVDQLEKEIELP 887
             S N++DYH+I+D++P ++ +YF  +  D + LS  Q ++LL +GLQ+K++D LEKE  LP
Sbjct:   831 SNNLLDYHVIVDLLPKLAHLYFSGKFPDSVKLSPVQQSVLLALGLQYKTIDTLEKEFNLP 890

Query:   888 SGQLMGLFNRIIRKVVKLFNEVQEKAIEEQMVAAKDVVME--------PTMKTLSDDLDE 939
             S QL+ +  ++ +K++K  +E++ K IEE++ + K             P  ++L ++L E
Sbjct:   891 SNQLLAMLVKLSKKIMKCIDEIETKDIEEELGSNKKTESSNSKLPEFTPLQQSLEEELQE 950

Query:   940 AAKEFQ-EKHKKEVGKLKSMDLSEYIIRGDDEEWNEVLNKAGPNASIISLKSDKKRKLEA 998
              A E      +K+   + ++DL +Y IRG++E+W     KA  N  I         R +
Sbjct:   951 GADEAMLALREKQRELINAIDLEKYAIRGNEEDW-----KAAEN-QIQKTNGKGARVVSI 1004

Query:   999 KQEPKQSKKL--KNRETKNKKDMKLKRKK 1025
             K E +++  L   +++TK K   K K +K
Sbjct:  1005 KGEKRKNNSLDASDKKTKEKPSSKKKFRK 1033
```

Pedant information for DKFZphtes3_6c11, frame 3

Report for DKFZphtes3_6c11.3

```
[LENGTH]      1025
[MW]          115704.57
[pI]          8.50
[HOMOL]       PIR:S55151 probable membrane protein YNL132w - yeast (Saccharomyces cerevisiae)
0.0
[FUNCAT]      10.99 other signal-transduction activities  [S. cerevisiae, YNL132w] 0.0
[FUNCAT]      r general function prediction       [H. influenzae, HI1254] 2e-05
[PROSITE]     ATP_GTP_A      1
[PROSITE]     RGD    1
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY    11.80 %
```

```
SEQ    MHRKKVDNRIRILIENGVAERQRSLFVVVGDRGKDQVVILHHMLSKATVKARPSVLWCYK
SEG    ............................................................
PRD    ccccccchhhhhhccccccceeeeeeeeecccccceeeeehhhhhhhhhhccceeehhhh

SEQ    KELGFSSHRKKRMRQLQKKIKNGTLNIKQDDPFELFIAATNIRYCYYNETHKILGNTFGM
SEG    ............................................................
PRD    hhhccccchhhhhhhhhhhhhhhhccccccccccceeeeeccccceeeeeccccceeeccccee

SEQ    CVLQDFEALTPNLLARTVETVEGGGLVVILLRTMNSLKQLYTVTMDVHSRYRTEAHQDVV
SEG    ................xxxxxxxxxxxxxxx.............................
PRD    eehhhhhccccchhhhhhhhhccccceeeeeeecchhhhhhhhhhhhhhhhhhhhhhhhh

SEQ    GRFNERFILSLASCKKCLVIDDQLNILPISSHVATMEALPPQTPDESLGPSDLELRELKE
SEG    ............................................................
PRD    hhhhhhhhhhhccccceeeeeeecceeeeecccccccccccccccccccchhhhhhhhh

SEQ    SLQDTQPVGVLVDCCKTLDQAKAVLKFIEGISEKTLRSTVALTAARGRGKSAALGLAIAG
SEG    ...................................................xxxxxxxxx
PRD    hhccccccceeeeehhhhhhhhhhhhhhhhhhhhhhhhhhheeeccccccchhhhhhhhhh

SEQ    AVAFGYSNIFVTSPSPDNLHTLFEFVFKGFDALQYQEHLDYEIIQSLNPEFNKAVIRVNV
SEG    xxx.........................................................
PRD    hhhhcccceeecccccccchhhhhhhhhhhhhhhhhhhheeeeecccccccceeeeeeh

SEQ    FREHRQTIQYIHPADAVKLGQAELVVIDEAAAIPLPLVKSLLGPYLVFMASTINGYEGTG
SEG    ............................................................
PRD    hhhhhhheeeeccccccccccceeeehhhhhcchhhhhhhcccceeeeeeeeccccccccc

SEQ    RSLSLKLIQQLRQQSAQSQVSTTAENKTTTTARLASARTLHEVSLQESIRYAPGDAVEKW
SEG    ........xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx....................
PRD    cchhhhhhhhhhhhhhhhhhhhhcccccccccchhhhhhhhhhhhhhceeeccccchhhh

SEQ    LNDLLCLDCLNITRIVSGCPLPEACELYYVNRDTLFCYHKASEVFLQRLMALYVASHYKN
SEG    xxxxxxxxxxx.................................................
PRD    hhhhhhcccccceeeccccccccceeeeeeecccccccccchhhhhhhhhhhhhhhhhhccc

SEQ    SPNDLQMLSDAPAHHLFCLLPPVPPTQNALPEVLAVIQVCLEGEISRQSILNSLSRGKKA
SEG    ............................................................
PRD    cccccccccccccceeeeeeccccccccccccchhhhhhhhhhcccccchhhhhhhhcccccc

SEQ    SGDLIPWTVSEQFQDPDFGGLSGGRVVRIAVHPDYQGMGYGSRALQLLQMYYEGRFPCLE
SEG    ............................................................
PRD    ccccchhhhhhhhhhhcccccccccceeeeeeecccccccccccchhhhhhhhhhhhccccchhh

SEQ    EKVLETPQEIHTVSSEAVSLLEEVITPRKDLPPLLLKLNERPAERLDYLGVSYGLTPRLL
```

```
SEG     ...........................xxxxxxxxxx.......................
PRD     hhhhcccccchhhhhhhhhhhhhhccccccccccccccccccceeeeccccccccchhhh

SEQ     KFWKRAGFVPVYLRQTPNDLTGEHSCIMLKTLTDEDEADQGGWLAAFWKDFRRRFLALLS
SEG     ............................................................
PRD     hhhhcccceeeeeccccccccceeeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhh

SEQ     YQFSTFSPSLALNIIQNRNMGKPAQPALSREELEALFLPYDLKRLEMYSRNMVDYHLIMD
SEG     ............................................................
PRD     hhhhcchhhhhhhhhhhcccccccchhhhhhhhhhhhccchhhhhhhhhccchhhhhhhhh

SEQ     MIPAISRIYFLNQLGDLALSAAQSALLLGIGLQHKSVDQLEKEIELPSGQLMGLFNRIIR
SEG     ............xxxxxxxxxxxxxxxxxxxx............................
PRD     hhhhhhhhhhhcccchhhhhhhhhhhhhhcchhhhhhhhhhhhhhccccchhhhhhhhhh

SEQ     KVVKLFNEVQEKAIEEQMVAAKDVVMEPTMKTLSDDLDEAAKEFQEKHKKEVGKLKSMDL
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhhhhhcc

SEQ     SEYIIRGDDEEWNEVLNKAGPNASIISLKSDKKRKLEAKQEPKQSKKLKNRETKNKKDMK
SEG     .............................................xxxxxxxxxxxxxxx
PRD     cceeecccchhhhhhhhhcccccceeeeeeeccchhhhhhhhcccccccccccccccchhhh

SEQ     LKRKK
SEG     xxxxx
PRD     hhccc
```

Prosite for DKFZphtes3_6c11.3

PS00016	966->969	RGD	PDOC00016
PS00017	284->292	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphtes3_6c11.3)

DKFZphtes3_6d16
----

group: testes derived

DKFZphtes3_6d16 encodes a novel 695 amino acid protein nearly identical to a sequence from human PAC clone WUGSC:H_DJ1185I07.2.

The cDNA is different to the proposed gene model: it contains additional exons.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

WUGSC:H_DJ1185I07.2, differences to genmodel

differences to genmodel of WUGSC:H_DJ1185I07.2 two exons skippt,

Sequenced by BMFZ

Locus: /map="7q11.23-q21"

Insert length: 4572 bp
Poly A stretch at pos. 4540, polyadenylation signal at pos. 4520

```
   1 GGCGGCGCTA GCTTCGGAGT CTCCCGCGCG CACCTCAGCC GCCTCCTAGC
  51 GGCGCGGCGC TCGCTCCTAC GCCTAAAATG ACCAATGTGT GATTTCAGTG
 101 GAATAAATGG CGTCCAAAGT CACAGATGCT ATAGTCTGGT ATCAAAAGAA
 151 GATTGGAGCA TATGATCAAC AAATATGGGA AAAATCTGTT GAACAGAGAG
 201 AAATCAAGGG GCTAAGGAAT AAACCAAAGA AAACAGCACA TGTGAAACCA
 251 GACCTCATAG ATGTTGATCT TGTAAGAGGG TCTGCATTTG CAAAGGCAAA
 301 GCCTGAAAGT CCTTGGACTT CTCTGACCAG AAAGGGAATT GTTCGAGTTG
 351 TATTTTTCCC CTTTTTCTTC CGGTGGTGGT TACAAGTAAC ATCAAAGGTC
 401 ATCTTTTTCT GGCTTCTTGT CCTTTATCTT CTTCAAGTTG CTGCAATAGT
 451 ATTATTCTGC TCCACTTCTA GCCCACACAG CATACCTCTG ACAGAGGTGA
 501 TTGGGCCGAT ATGGCTGATG CTGCTCCTGG GAACTGTGCA TTGCCAGATT
 551 GTTTCCACAA GAACACCCAA ACCTCCTCTA AGTACAGGGG GTAAAAGAAG
 601 AAGGAAATTA AGAAAAGCAG CCCATTTGGA AGTACATAGG GAAGGAGATG
 651 GTTCTAGTAC CACAGATAAC ACACAAGAGG GAGCAGTTCA GAACCACGGT
 701 ACAAGCACCT CTCACAGCGT TGGCACTGTC TTCAGAGATC TCTGGCATGC
 751 TGCTTTCTTT TTATCAGGAT CAAAGAAAGC AAAGAATTCA ATTGATAAAT
 801 CAACTGAAAC TGACAATGGC TATGTATCCC TTGATGGGAA GAAGACTGTT
 851 AAAAGCGGTG AAGATGGAAT ACAAAACCAT GAACCTCAGT GTGAAACTAT
 901 TCGACCAGAA GAGACAGCCT GGAACACAGG AACACTGAGG AATGGTCCTA
 951 GCAAAGATAC CCAAAGGACA ATAACAAATG TCTCTGATGA AGTCTCCAGT
1001 GAGGAAGGTC CTGAAACAGG ATACTCATTA CGTCGTCATG TGGACAGGAC
1051 TTCTGAAGGT GTTCTTCGGA ATAGAAAGTC ACACCATTAT AAGAAACATT
1101 ACCCTAATGA GGACGCCCCT AAATCGGGTA CTAGTTGCAG CTCTCGCTGT
1151 TCAAGTTCCA GACAGGATTC TGAGAGTGCA AGGCCAGAAT CTGAAACAGA
1201 AGATGTGTTA TGGGAAGACT TGTTACATTG TGCAGAATGC CATTCATCTT
1251 GTACCAGTGA GACAGATGTG GAAAATCATC AGATTAATCC ATGTGTGAAA
1301 AAAGAATATA GAGATGACCC TTTTCATCAG AGTCATTTGC CCTGGCTCCA
1351 TAGTTCCCAC CCAGGATTAG AAAAAATAAG TGCTATAGTA TGGGAAGGTA
1401 ATGATTGTAA GAAAGCAGAC ATGTCTGTAC TTGAAATCAG TGGAATGATA
1451 ATGAACAGAG TGAACAGCCA TATACCAGGA ATAGGATACC AGATTTTTGG
1501 AAATGCAGTC TCTCTCATAC TGGGTTTAAC TCCATTTGTT TTCCGACTTT
1551 CTCAAGCTAC AGACTTGGAA CAACTCACAG CACATTCTGC TTCAGAACTT
1601 TATGTGATTG CATTTGGTTC TAATGAAGAT GTCATAGTTC TTTCTATGGT
1651 TATAATAAGT TTTGTGGTTC GCGTGTCTCT TGTGTGGATT TTCTTTTTTT
1701 TGCTCTGTGT AGCAGAAAGA ACTTATAAAC AGCGATTACT TTTTGCAAAA
1751 CTCTTTGGAC ATTTAACATC TGCAAGGAGG GCTCGAAAAT CTGAGGTTCC
1801 TCATTTCCGG TTGAAGAAAG TACAGAATAT AAAAATGTGG CTATCTCTCC
1851 GTTCCTATCT TAAGCGTCGA GGTCCTCAGC GATCAGTTGA TGTAATAGTT
1901 TCATCTGCTT TCTTATTGAC TATCTCAGTT GTATTTATCT GTTGTGCCCA
1951 GATAAACCTC TACTTGAAAA TGGAGAAAAA ACCTAACAAA AAGGAGGAAC
2001 TGACACTAGT GAATAATGTT TTAAAACTGG CTACTAAACT GCTAAAGGAG
2051 TTGGACAGTC CTTTTAGATT ATATGGGCTT ACAATGAATC CGCTGCTTTA
2101 TAACATCACC CAGGTTGTTA TCCTGTCAGC TGTTTCTGGT GTTATCAGTG
2151 ACTTGCTTGG ATTTAATTTA AAGCTATGGA AGATTAAGTC ATGACAATTC
2201 AAAGAAAAGA AGATGTAGCC TCTTTTCCAG AATAAGAGTA CTGACTAAGC
2251 TGCCTGAAAG CTTGTCACTG ATTCTTTGCT TCAGGAGTCT CAGCTAGGGA
2301 GTTGAAGTGT TTACATCAGA CTGTCTTGTG CAATTCTTAT ATTTATTTTA
2351 CTGGTTCACT TTTTTTTACA TTTATTTTAG TCTTTATATT TTTATTTTTA
2401 AGCATTGATG TACTTAGTTG TTGAAAGGGT GATGAAACTG ATATCCAGAT
2451 ACTTGAGATC CTGGTAATTG GTCATAAATA ATTGGCAAAA TAACAAATTG
2501 TGAAAATAGA AGCCATTGCT CAGCACCGTT TCTCCATCAA TGCCGTGAAC
2551 TTGCCTTACT TGAGGAAAAA TTCTTTAACT TTGGAATATT GCATTGAACT
2601 CAGCTATACA CATAAAACAT TTTCTTTGGT AAATCAAGAT CCAGTCAGGG
```

```
2651 TTTCTCTTGA ATTATTTTGG AACAATGCCA GGATCCAAAC TGATTAAGTT
2701 ACAGTTTAAG CACCCTTCAG TATTAATATA TACGGTATTA TATAACAGGT
2751 CAACAAGTGC TCTTTGATGA TAAAACTTGT AATAGAGCAA TAATTGTAAA
2801 TGGTTACCAT ACTGTAAGAT ATTTTGATAA AAATTAACTA GTAATACTTG
2851 TATTTATTTG AAACACTGGG CTGTTTGCAC AGCTCCAACT GTGCATGCTC
2901 AAAATGTGCA CTTTTTAAAA TTGTTACTTT TAATGCGTAT CTTTATATGG
2951 GATCTGTTAT AGTATACTAG GGCATGATAT GGTATCCTTT TGAGTGAGGT
3001 ATATACTCAT CTCACAAGTG AAGTGCCTAC TGATATTACT AAAGTACATT
3051 ATGTTTACTC AAGTAAATAA TTTTCTCCCC ATGGTACACT CTAGTGTAGG
3101 CTATTCATAC CACACTGAAA TGAACAACTG AAGAATAAGG CTAAGAACCA
3151 ATAAAATATT TCTCTAATTG CTAGTTGTAA AACTGTATCC AAATTTTCAG
3201 AAAAGACAGC TTCAGCTTGC AAATTCTATC CTCTAAACTT ATCTGGTGCA
3251 TTCTCCCCAC CCCACCCCCA TTATATAAGG GCTATTTTAG ATGCTTTTAA
3301 CCTCCCCAAC AAATAATTTG CCAAGTGTCC AATGAGAACT TATCATGTTG
3351 GTGTGTTAGG TAAATCGGGC AAATATGATA GTGTCTTACA TTGGGCCTTG
3401 ATTTTAAGTT GTTATATTTG TACAATCGAG TATTTTAGAA ATTACATGAA
3451 ACATGAAACA GTTTTTGCAA TTTTTTTTAA ACTGGGCATC TGGTTTCTAA
3501 AAATTTATTT GAAACAATCT AGAATTTTCT TGGTGCAAAG TGTATCATGT
3551 GGAATATCCT CATATTTTTA CCATATTTTA AGAACTTTAA GACGATTAAT
3601 TGTAAATAAT TTATTTGATT GGTGCAGTTC TAATCCCTAA ATCATAATCT
3651 TAAAATCAGG AATGTGTGGA GAACAGAGCC ATGTCATATC ACTTTGCTCT
3701 TACCATTCCT TTTGATCAGC CTCAATTCAG CCTCATTGTG TAGTATGTTT
3751 TTTCTTTCTA TGAAAAACAA CAGAAAGCAT TTCATTTTAT TTGCCTATGT
3801 TCAAATATGT TTAATAATGA CCAAAGTGCA TTCTGAGTTT TTTCAAGGAA
3851 TGTAATACTG GAGCTTTAAG AACATACTTA GTTTCTCATG TGAAAACTTA
3901 GGCTTTGTCT GATGTTTTTC CTTCCTCTAT TGTCTAATGT TGAGGTTGTT
3951 TTTAGGAATT ATGTTTTATA AACTTTTTCA ATATAAGGTA CATGCCTATA
4001 CAGAACTTAA CATTTTGCAC AGAATATATC AAATATATTT TGAGAAAAAA
4051 AGTACGGCAT GAGTTCTGTT AGGAATAAAA GATGAAACTA TTGTATCTCA
4101 CAAAAAATCT TATTTCAGAA TGGAAATATT TTTGAGAAAA GTAGCTGAGT
4151 ATACTGGTTT AAGAAAATGC TTGTTTTAGA TTGAGGTTAA CTTAGAGTTG
4201 GGAGTTGATT TATTAAGTAC AGTATACCTC TCAACAGTTT ATAAATAATA
4251 TGTTGAATTA TGTCAGTGTG GGCAGCAGTA GAATACTAAA AGGAAAATGT
4301 CATGTTAAGC AATTTCAGAA CATTAACTGA ACTATTTTCA AAGCAGAAAA
4351 ATTGACATTG CTGCCTTTAA GAATACCATG AATGTAAGAA ATTGAAAGAA
4401 ATTGTAAAAT ATCACATAAT ATAGAAATGG CAGTTCAAAG AGAATTGTGG
4451 CAGATGTTGT GTGTGAACTG TTGTTTCTTT GCCACATGTG TTGTATTTGA
4501 AAGTTTTACA GTAAGTTTAA AATAAAACAT TCTGTGACTG AAAAAAAAAA
4551 AAAAAAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 107 bp to 2191 bp; peptide length: 695
Category: known protein
Classification: unclassified
Prosite motifs: CYTOCHROME_C (375-381)

```
  1 MASKVTDAIV WYQKKIGAYD QQIWEKSVEQ REIKGLRNKP KKTAHVKPDL
 51 IDVDLVRGSA FAKAKPESPW TSLTRKGIVR VVFFPFFFRW WLQVTSKVIF
101 FWLLVLYLLQ VAAIVLFCST SSPHSIPLTE VIGPIWLMLL LGTVHCQIVS
151 TRTPKPPLST GGKRRRKLRK AAHLEVHREG DGSSTTDNTQ EGAVQNHGTS
201 TSHSVGTVFR DLWHAAFFLS GSKKAKNSID KSTETDNGYV SLDGKKTVKS
251 GEDGIQNHEP QCETIRPEET AWNTGTLRNG PSKDTQRTIT NVSDEVSSEE
301 GPETGYSLRR HVDRTSEGVL RNRKSHHYKK HYPNEDAPKS GTSCSSRCSS
351 SRQDSESARP ESETEDVLWE DLLHCAECHS SCTSETDVEN HQINPCVKKE
401 YRDDPFHQSH LPWLHSSHPG LEKISAIVWE GNDCKKADMS VLEISGMIMN
451 RVNSHIPGIG YQIFGNAVSL ILGLTPFVFR LSQATDLEQL TAHSASELYV
501 IAFGSNEDVI VLSMVIISFV VRVSLVWIFF FLLCVAERTY KQRLLFAKLF
551 GHLTSARRAR KSEVPHFRLK KVQNIKMWLS LRSYLKRRGP QRSVDVIVSS
601 AFLLTISVVF ICCAQINLYL KMEKKPNKKE ELTLVNNVLK LATKLLKELD
651 SPFRLYGLTM NPLLYNITQV VILSAVSGVI SDLLGFNLKL WKIKS
```

BLASTP hits

```
No BLASTP hits available

            Alert BLASTP hits for DKFZphtes3_6d16, frame 2

PIR:S38170 SRP40 protein - yeast (Saccharomyces cerevisiae), N = 1,
Score = 100, P = 0.08

TREMBL:AC004990_1 gene: "WUGSC:H_DJ1185I07.2";      Homo sapiens PAC clone
DJ1185I07 from 7q11.23-q21, complete sequence., N = 2, Score = 2693, P
= 0


>TREMBL:AC004990_1 gene: "WUGSC:H_DJ1185I07.2"; Homo sapiens PAC clone
     DJ1185I07 from 7q11.23-q21, complete sequence.
            Length = 588

  HSPs:

 Score = 2693 (404.1 bits), Expect = 0.0e+00, Sum P(2) = 0.0e+00
 Identities = 510/515 (99%), Positives = 512/515 (99%)

Query:    35 GLRNKPKKTAHVKPDLIDVDLVRGSAFAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQV 94
             GLRNKPKKTAHVKPDLIDVDLVRGSAFAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQV
Sbjct:    1 GLRNKPKKTAHVKPDLIDVDLVRGSAFAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQV 60

Query:    95 TSKVIFFWLLVLYLLQVAAIVLFCSTSSPHSIPLTEVIGPIWLMLLLGTVHCQIVSTRTP 154
             TSKVIFFWLLVLYLLQVAAIVLFCSTSSPHSIPLTEVIGPIWLMLLLGTVHCQIVSTRTP
Sbjct:   61 TSKVIFFWLLVLYLLQVAAIVLFCSTSSPHSIPLTEVIGPIWLMLLLGTVHCQIVSTRTP 120

Query:   155 KPPLSTGGKRRRKLRKAAHLEVHREGDGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWH 214
             KPPLSTGGKRRRKLRKAAHLEVHREGDGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWH
Sbjct:   121 KPPLSTGGKRRRKLRKAAHLEVHREGDGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWH 180

Query:   215 AAFFLSGSKKAKNSIDKSTETDNGYVSLDGKKTVKSGEDGIQNHEPQCETIRPEETAWNT 274
             AAFFLSGSKKAKNSIDKSTETDNGYVSLDGKKTVKSGEDGIQNHEPQCETIRPEETAWNT
Sbjct:   181 AAFFLSGSKKAKNSIDKSTETDNGYVSLDGKKTVKSGEDGIQNHEPQCETIRPEETAWNT 240

Query:   275 GTLRNGPSKDTQRTITNVSDEVSSEEGPETGYSLRRHVDRTSEGVLRNRKSHHYKKHYPN 334
             GTLRNGPSKDTQRTITNVSDEVSSEEGPETGYSLRRHVDRTSEGVLRNRKSHHYKKHYPN
Sbjct:   241 GTLRNGPSKDTQRTITNVSDEVSSEEGPETGYSLRRHVDRTSEGVLRNRKSHHYKKHYPN 300

Query:   335 EDAPKSGTSCSSRCSSSRQDSESARPESETEDVLWEDLLHCAECHSSCTSETDVENHQIN 394
             EDAPKSGTSCSSRCSSSRQDSESARPESETEDVLWEDLLHCAECHSSCTSETDVENHQIN
Sbjct:   301 EDAPKSGTSCSSRCSSSRQDSESARPESETEDVLWEDLLHCAECHSSCTSETDVENHQIN 360

Query:   395 PCVKKEYRDDPFHQSHLPWLHSSHPGLEKISAIVWEGNDCKKADMSVLEISGMIMNRVNS 454
             PCVKKEYRDDPFHQSHLPWLHSSHPGLEKISAIVWEGNDCKKADMSVLEISGMIMNRVNS
Sbjct:   361 PCVKKEYRDDPFHQSHLPWLHSSHPGLEKISAIVWEGNDCKKADMSVLEISGMIMNRVNS 420

Query:   455 HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSNEDVIVLSM 514
             HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSNEDVIVLSM
Sbjct:   421 HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSNEDVIVLSM 480

Query:   515 VIISFVVRVSLVWIFFFLLCVAERTYKQRLLFAKL 549
             VIISFVVRVSLVWIFFFLLCVAERTYKQ  L+ K+
Sbjct:   481 VIISFVVRVSLVWIFFFLLCVAERTYKQINLYLKM 515

 Score = 409 (61.4 bits), Expect = 0.0e+00, Sum P(2) = 0.0e+00
 Identities = 92/115 (80%), Positives = 98/115 (85%)

Query:   595 DVIVSS----AFLLTISVVFI-----CCA-----QINLYLKMEKKPNKKEELTLVNNVLK 640
             DVIV S    +F++ +S+V+I     C A     QINLYLKMEKKPNKKEELTLVNNVLK
Sbjct:   474 DVIVLSMVIISFVVRVSLVWIFFFLLCVAERTYKQINLYLKMEKKPNKKEELTLVNNVLK 533

Query:   641 LATKLLKELDSPFRLYGLTMNPLLYNITQVVILSAVSGVISDLLGFNLKLWKIKS 695
             LATKLLKELDSPFRLYGLTMNPLLYNITQVVILSAVSGVISDLLGFNLKLWKIKS
Sbjct:   534 LATKLLKELDSPFRLYGLTMNPLLYNITQVVILSAVSGVISDLLGFNLKLWKIKS 588


           Pedant information for DKFZphtes3_6d16, frame 2
           -----------------------------------------------

                    Report for DKFZphtes3_6d16.2


[LENGTH]        695
[MW]            78466.68
[pI]            9.30
[HOMOL]         TREMBL:AC004990_1 gene: "WUGSC:H_DJ1185I07.2"; Homo sapiens PAC clone DJ1185I07
from 7q11.23-q21, complete sequence. 0.0
```

```
[PROSITE]      CYTOCHROME_C     1
[KW]           TRANSMEMBRANE    6
[KW]           LOW_COMPLEXITY      5.32 %

SEQ   MASKVTDAIVWYQKKIGAYDQQIWEKSVEQREIKGLRNKPKKTAHVKPDLIDVDLVRGSA
SEG   ............................................................
PRD   ccceeeeehhhhhhhcccchhhhhhhhhhhhhhhcccccccccccccccceeeeeeccch
MEM   ............................................................

SEQ   FAKAKPESPWTSLTRKGIVRVVFFPFFFRWWLQVTSKVIFFWLLVLYLLQVAAIVLFCST
SEG   ..................xxxxxxxxxxx...............................
PRD   hhhhcccccccccccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhhhhhheeecc
MEM   ......................MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM........

SEQ   SSPHSIPLTEVIGPIWLMLLLGTVHCQIVSTRTPKPPLSTGGKRRRKLRKAAHLEVHREG
SEG   ..........................................xxxxxxxx..........
PRD   ccccccceeeeehhhhhhhhhhhhheeeeeecccccccccccchhhhhhhhhhhheeecc
MEM   ......MMMMMMMMMMMMMMMMMMMMMMM...............................

SEQ   DGSSTTDNTQEGAVQNHGTSTSHSVGTVFRDLWHAAFFLSGSKKAKNSIDKSTETDNGYV
SEG   ............................................................
PRD   cccccccccceeeeeeecccccccchhhhhhhhhhhhhcccchhhhhccccccccccccc
MEM   ............................................................

SEQ   SLDGKKTVKSGEDGIQNHEPQCETIRPEETAWNTGTLRNGPSKDTQRTITNVSDEVSSEE
SEG   ............................................................
PRD   cccccceeeccccccccccccccccccccceeeecccccccccccccceeeeccccccccc
MEM   ............................................................

SEQ   GPETGYSLRRHVDRTSEGVLRNRKSHHYKKHYPNEDAPKSGTSCSSRCSSSRQDSESARP
SEG   .......................................xxxxxxxxxxxxxxxxxx...
PRD   ccccceeeeeecccccccchhhhhhcccccccccccccccccccccccccccccccccccc
MEM   ............................................................

SEQ   ESETEDVLWEDLLHCAECHSSCTSETDVENHQINPCVKKEYRDDPFHQSHLPWLHSSHPG
SEG   ............................................................
PRD   cccchhhhhhhhhhhcccccccccccccccccccccceeeeecccccccccccccccccccc
MEM   ............................................................

SEQ   LEKISAIVWEGNDCKKADMSVLEISGMIMNRVNSHIPGIGYQIFGNAVSLILGLTPFVFR
SEG   ............................................................
PRD   cccceeeeeecccccccceeeeehhhhhhhcccccccccccccccccceeeccccccchh
MEM   ...............................MMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ   LSQATDLEQLTAHSASELYVIAFGSNEDVIVLSMVIISFVVRVSLVWIFFFLLCVAERTY
SEG   ............................................................
PRD   hhhhhhhhhhhcccceeeeeeeecccccceeeehhhhhhhcchhhhhhhhhhhhhhhhhhh
MEM   ......................MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM........

SEQ   KQRLLFAKLFGHLTSARRARKSEVPHFRLKKVQNIKMWLSLRSYLKRRGPQRSVDVIVSS
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhcccccceeeeeehhhhhhhhhhhhccccceeeeeeeee
MEM   .....................................................MMMMMMM

SEQ   AFLLTISVVFICCAQINLYLKMEKKPNKKEELTLVNNVLKLATKLLKELDSPFRLYGLTM
SEG   ............................................................
PRD   eeeeeeeeeeeeeehhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhccccceeeeccc
MEM   MMMMMMMMMMMMMMMMMMM.........................................

SEQ   NPLLYNITQVVILSAVSGVISDLLGFNLKLWKIKS
SEG   ...................................
PRD   cchhhheeeeeeeeeecchhhhccceeeeeeeccc
MEM   MMMMMMMMMMMMMMMMMMMMMMMMM..........
```

Prosite for DKFZphtes3_6d16.2

```
PS00190     375->381    CYTOCHROME_C        PDOC00169
```

(No Pfam data available for DKFZphtes3_6d16.2)

DKFZphtes3_72k11
----

group: testes derived
DKFZphtes3_72k11 encodes a novel 233 amino acid protein with similarity to S.pombe hypothetical repeat-containing protein.

The novel protein contains 5 leucine zippers and a microbodies C-terminal targeting signal (S-K-L) signature. This sequence is responsible for transport of proteins from free polysomes into the microbodies.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to S.pombe hypothetical repeat-containing protein

complete cDNA, complete cds, 6 EST hits (3 from testis derived librarys)

Sequenced by DKFZ

Locus: unknown

Insert length: 1134 bp
Poly A stretch at pos. 1124, polyadenylation signal at pos. 1088

```
   1 AACCTTTCAA GTGCCCCCTC CTTTCCTTAA AGTCTTTTAT AGGGGTCCCC
  51 TTCTTGGCCA TCTCCATCCT GTGAGTCAGG ACTGAAAGGG CACAGACAGG
 101 TCACTGCCAG CATTGTTGGG GCAAGCCTGC AAGCACGCAT CACTGGGGAT
 151 CTGACATGAC AATGGCCGCC TGCCCCCTCT GAGGGCTACA GGACTTACCC
 201 CAGTGGGAAG CAGCTAAGCA GGTCTGACCA GCCGACCTGG ACCTGGCCAA
 251 GGGTCCTGTC ATCCCTCATG GCCACCCCGC CATTCCGGCT GATAAGGAAG
 301 ATGTTTTCCT TCAAGGTGAG CAGATGGATG GGGCTTGCCT GCTTCCGGTC
 351 CCTGGCGGCA TCCTCTCCCA GTATTCGCCA GAAGAAACTA ATGCACAAGC
 401 TGCAGGAGGA AAAGGCTTTT CGCGAAGAGA TGAAAATTTT TCGTGAAAAA
 451 ATAGAGGACT TCAGGGAAGA GATGTGGACT TTCCGAGGCA AGATCCATGC
 501 TTTCCGGGGC CAGATCCTGG GTTTTTGGGA AGAGGAGAGA CCTTTCTGGG
 551 AAGAGGAGAA AACCTTCTGG AAAGAGGAAA AATCCTTCTG GGAAATGGAA
 601 AAGTCTTTCA GGGAGGAAGA GAAAACTTTC TGGAAAAAGT ACCGCACTTT
 651 CTGGAAGGAG GATAAGGCCT TCTGGAAAGA GGACAATGCC TTATGGGAAA
 701 GAGACCGGAA CCTTCTTCAG GAGGACAAGG CCCTGTGGGA GGAAGAAAAG
 751 GCCCTGTGGG TAGAGGAAAG AGCCCTCCTT GAGGGGGAGA AAGCCCTGTG
 801 GGAAGATAAA ACGTCCCTCT GGGAGGAAGA GAATGCCCTC TGGGAGGAAG
 851 AGAGGGCCTT CTGGATGGAG AACAATGGCC ACGTTGCCGG AGAGCAGATG
 901 CTCGAAGATG GGCCCCACAA CGCCAACAGA GGGCAGCGCT TGCTGGCCTT
 951 CTCCCGAGGC AGGGCGTAGC CAGCATGCAG GTGCAGGGCC CTGTGGTCCA
1001 GACTCCCCTG GGTTGGGATT CAAGTCCAGG GTGAGCCCAT GTGCTGGAGA
1051 AAATACACAC TCATTGGTCT CCTTGCTTTG AAAGATCCAA TAAAGTCCTG
1101 AGGCAAGGTT TGGAAAACCA ACTTAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 268 bp to 966 bp; peptide length: 233
Category: similarity to known protein
Prosite motifs: MICROBODIES_CTER (231-234)
LEUCINE_ZIPPER (142-164)
LEUCINE_ZIPPER (149-171)
LEUCINE_ZIPPER (156-178)
LEUCINE_ZIPPER (163-185)
LEUCINE_ZIPPER (170-192)
LEUCINE_ZIPPER (170-192)

```
  1 MATPPFRLIR KMFSFKVSRW MGLACFRSLA ASSPSIRQKK LMHKLQEEKA
 51 FREEMKIFRE KIEDFREEMW TFRGKIHAFR GQILGFWEEE RPFWEEEKTF
101 WKEEKSFWEM EKSFREEEKT FWKKYRTFWK EDKAFWKEDN ALWERDRNLL
151 QEDKALWEEE KALWVEERAL LEGEKALWED KTSLWEEENA LWEEERAFWM
201 ENNGHVAGEQ MLEDGPHNAN RGQRLLAFSR GRA
```

BLASTP hits

Entry SPCC330_4 from database TREMBLNEW:
gene: "SPCC330.04c"; product: "hypothetical repeat-containing protein";
S.pombe chromosome III cosmid c330.
Score = 149, P = 1.6e-08, identities = 55/187, positives = 88/187

Entry A45973 from database PIR:
trichohyalin - human
Score = 147, P = 3.0e-07, identities = 57/194, positives = 94/194

Alert BLASTP hits for DKFZphtes3_72k11, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_72k11, frame 1

Report for DKFZphtes3_72k11.1

[LENGTH]	233	
[MW]	28752.65	
[pI]	5.70	
[PROSITE]	LEUCINE_ZIPPER 5	
[PROSITE]	MICROBODIES_CTER	1
[PROSITE]	MYRISTYL 1	
[PROSITE]	CK2_PHOSPHO_SITE	3
[PROSITE]	PKC_PHOSPHO_SITE	4
[KW]	All_Alpha	
[KW]	LOW_COMPLEXITY	15.45 %

```
SEQ  MATPPFRLIRKMFSFKVSRWMGLACFRSLAASSPSIRQKKLMHKLQEEKAFREEMKIFRE
SEG  ............................................................
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhh

SEQ  KIEDFREEMWTFRGKIHAFRGQILGFWEEERPFWEEEKTFWKEEKSFWEMEKSFREEEKT
SEG  .........................xxxxxxxxxxxxxxxxxxxxxxxx...........
PRD  hhhhhhhhhhhhhhhhcccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhh

SEQ  FWKKYRTFWKEDKAFWKEDNALWERDRNLLQEDKALWEEEKALWVEERALLEGEKALWED
SEG  ............................................................
PRD  hhhhcccccccccchhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  KTSLWEEENALWEEERAFWMENNGHVAGEQMLEDGPHNANRGQRLLAFSRGRA
SEG  ...xxxxxxxxxxxx......................................
PRD  ccchhhhhhhhhhhhhhhhhhcccccchhhhhhcccccccccchhhhhhhhccc
```

Prosite for DKFZphtes3_72k11.1

PS00005	14->17	PKC_PHOSPHO_SITE	PDOC00005
PS00005	35->38	PKC_PHOSPHO_SITE	PDOC00005
PS00005	71->74	PKC_PHOSPHO_SITE	PDOC00005
PS00005	113->116	PKC_PHOSPHO_SITE	PDOC00005
PS00006	106->110	CK2_PHOSPHO_SITE	PDOC00006
PS00006	113->117	CK2_PHOSPHO_SITE	PDOC00006
PS00006	183->187	CK2_PHOSPHO_SITE	PDOC00006
PS00008	81->87	MYRISTYL	PDOC00008
PS00342	231->234	MICROBODIES_CTER	PDOC00299
PS00029	142->164	LEUCINE_ZIPPER	PDOC00029
PS00029	149->171	LEUCINE_ZIPPER	PDOC00029
PS00029	156->178	LEUCINE_ZIPPER	PDOC00029
PS00029	163->185	LEUCINE_ZIPPER	PDOC00029
PS00029	170->192	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_72k11.1)

DKFZphtes3_72k15
----

group: cell structure and motility

DKFZphtes3_72k15 encodes a novel 188 amino acid protein with strong similarity to Rattus norvegicus actin-filament binding protein Frabin.

FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin-binding protein. The gene locus fgd1 seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an esin yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus.
The novel protein seems to be the human orthologue of rat frabin.

The new protein can find application in modulating of cell structure and motility as well as modulation of the JNK/SAPK pathway.

strong similarity to actin-filament binding protein Frabin

2 EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1845 bp
Poly A stretch at pos. 1835, polyadenylation signal at pos. 1816

```
   1 GTGATGGAGA GTGCTGTTAT GATAGATGAA TCTAGGAAAG CCTCTTTGGA
  51 GATGTGATAC CTGAACAGAA CCCCGAATGA TAAGAAGAAA TACCAGTGTT
 101 TTAGGAGAGA TTGTCCTAAG CAGAGAACAG CAGCTGCAAA GACCCCAAGA
 151 CACATACACT TGGTTATTAA GAATGGGAGC AGCAAGGAGT ATGGCAAGAA
 201 CACAGTGAGT TTTCCCTTGA GTGTGTGAGG AAGCCCTCAG AGTTTGTGAC
 251 TGACTTGTAG AGGTTCTAGT GGAGGGGATC AGAGTGGAAA CAAAGAGACC
 301 AGTTAAAAAG GTATGGCAGC ATGAATAAAA AAGTTTTGAG AGTATTCATT
 351 ATGCCTTCCA AATAAAAAAC TCTTTGGTTC ATAATTTGTT CATAAATTAA
 401 GGACTGGCTA CACTGTACTA TTTAAAAATG TTAAGAAACA TCAATAAGTA
 451 AAAATGTTAG GAAGAGATGA TAAATACGTA AGTATTATAT CTAACTAAGT
 501 CTTTACTAAC TAGTCACATT ATTAAACAGT GCAAGGATCA AGAAAAGTTA
 551 AGCGTTGAAA AATAAATAAA TAAGTTATAA ATAAAATAAA CAGCCCAAGG
 601 AAATGTTCCA GTCCCCATAG GTAGACTCGG GGTCATCTTC TTTATTTAAA
 651 TCTTTATTTA AATGTGGATA GCATCCCAAG AGACTTGGGT CTACACTAAG
 701 AATATTCAAA TCCATGTTTC TGAAACCATC AGAGATAGAA AAAAAAAGTA
 751 GCGAATATCC CTTTTCAACT GGAATAAACT TGTCTTAATT CTAGAACTTT
 801 TCCATACCAA TGTTTTCATG CTTCCTTTGT ATTTTATCTT TTAGCTCATT
 851 ATCAAATTAT AGTGATTTGA AGAAAGAGTC TGCTGTGAAC CTAAATGCTC
 901 CTAGAACCCC AGGAAGGCAT GGATTGACAA CCACACCTCA ACAAAAACTC
 951 CTCTCCCAGC ACTTGCCACA GAGGCAGGGA AATGATACAG ATAAGACTCA
1001 GGGTGCACAG ACTTGTGTGG CCAACGGTGT AATGGCAGCA CAAAACCAGA
1051 TGGAATGTGA GGAGGAGAAA GCTGCCACTC TTAGCTCAGA TACTTCTATT
1101 CAAGCTTCTG AACCCTTGCT TGATACGCAC ATAGTGAATG GAGAAAGAGA
1151 TGAAACTGCC ACAGCTCCTG CATCACCCAC AACAGATAGC TGTGATGGAA
1201 ATGCTTCTGA CAGTAGCTAC AGGACTCCAG GCATAGGCCC AGTGCTCCCC
1251 CTAGAAGAAA GAGGGGCAGA AACAGAAACC AAGGTACAAG AGAGGGAAAA
1301 TGGGGAAAGC CCTCTGGAAC TGGAGCAGCT GGACCAGCAC CATGAGATGA
1351 AGGTAGAGCA TGAGACTAGC TCATGAGCAG GGAAAACCCT GCCTATTCGA
1401 TTGTTGTCTT AAAACTCTTT ATTTATTGCA CCCCTGAAAT GTATGAATCA
1451 GATCACCCAC ACTGGCAGTT AAACGATTTT CAAGCTCTGG CTGCTGATTA
1501 GCATTTCCCC TATGCTCTAA GCAGATATTT CACTTTTTCT TTTCATGTAG
1551 TTTCTGTTAA TATCTCTGTT GTAATTTCAG GAGTCAGAAC AGTGTGGAAA
1601 CTTTAATATA GGAAATCCAC AAATGTATTG TTTTTACATA GAAAGAAAAT
1651 GTTCCTTGTT GCTCTAGATG TTGGTGCTGT ATCCCTAATA CTTACGGGCC
1701 AAGCAAGAAG AAATTGTATA ATCTTTGTTG TTCAGAAGTT TCTAATAGAA
1751 TAAATAGGCC TGTAAGATGA ACTTGCCACT AGTAAATGTT ACTTTTAAGG
1801 ACATGAATAT GGAAGTATTA AATTATTCAA CAGATAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

98334590:
Frabin, a novel FGD1-related actin filament-binding protein capable of changing cell shape
and activating c-Jun N-terminal kinase.

Peptide information for frame 3

ORF from 810 bp to 1373 bp; peptide length: 188
Category: similarity to known protein
Classification: Cell structure/motility

```
  1 MFSCFLCILS FSSLSNYSDL KKESAVNLNA PRTPGRHGLT TTPQQKLLSQ
 51 HLPQRQGNDT DKTQGAQTCV ANGVMAAQNQ MECEEEKAAT LSSDTSIQAS
101 EPLLDTHIVN GERDETATAP ASPTTDSCDG NASDSSYRTP GIGPVLPLEE
151 RGAETETKVQ ERENGESPLE LEQLDQHHEM KVEHETSS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_72k15, frame 3

TREMBL:AF038388_1 product: "actin-filament binding protein Frabin"; Rattus norvegicus actin-filament binding protein Frabin mRNA, complete cds., N = 1, Score = 428, P = 1.8e-39

>TREMBL:AF038388_1 product: "actin-filament binding protein Frabin"; Rattus norvegicus actin-filament binding protein Frabin mRNA, complete cds.
Length = 766

HSPs:

Score = 428 (64.2 bits), Expect = 1.8e-39, P = 1.8e-39
Identities = 90/174 (51%), Positives = 115/174 (66%)

```
Query:    12 SSLSNYSDLKKESAVNLNAPRTPGRHGLTTTPQQKLLSQHLPQRQGNDTDKTQGAQTCVA 71
             S LS+Y+D++K+S +NLN P+TP +HGLT+T  QKL S   PQ+Q  D+D+ QG   C+A
Sbjct:    31 SVLSSYTDVQKDSTMNLNIPQTPRQHGLTSTTPQKLPSHKSPQKQEKDSDQNQGQHGCLA 90

Query:    72 NGVMAAQNQMECEEEKAATLSSDTSIQASEPLLDTHIVNGERDETATAPASPTTDSCDGN 131
             NGV AAQ+QMECE EK A LS +T  Q +      D H++NG R+ET T  AS  T+S D N
Sbjct:    91 NGVAAAQSQMECETEKEAALSPETDTQTAAASPDAHVLNGVRNETTTDSASSVTNSHDEN 150

Query:   132 ASDSSYRTPGIGPVLPLEERGAETETKVQERENGESPLELEQLDQHHEMKVEHE 185
             A DSS RT G     LP +E       E ++QERENG S    L  LDQHHE+K   +E
Sbjct:   151 ACDSSCRTQGTDLGLPSKEGEPVIEAELQERENGLSTEGLNPLDQHHEVKETNE 204
```

Pedant information for DKFZphtes3_72k15, frame 3

Report for DKFZphtes3_72k15.3

[LENGTH] 188
[MW] 20388.32
[pI] 4.62
[HOMOL] TREMBL:AF038388_1 product: "actin-filament binding protein Frabin"; Rattus norvegicus actin-filament binding protein Frabin mRNA, complete cds. 2e-38
[KW] All_Alpha
[KW] SIGNAL_PEPTIDE 16
[KW] LOW_COMPLEXITY 12.77 %

```
SEQ   MFSCFLCILSFSSLSNYSDLKKESAVNLNAPRTPGRHGLTTTPQQKLLSQHLPQRQGNDT
SEG   .xxxxxxxxxxxxxx.............................................
PRD   ccchhhhhcccccccccccccccccccccccccccccccccccchhhhhhhccccccccc

SEQ   DKTQGAQTCVANGVMAAQNQMECEEEKAATLSSDTSIQASEPLLDTHIVNGERDETATAP
```

```
SEG       .......................................................xxxxx
PRD       ccccccceeecchhhhhhhhhhhhhhhhhhhhccccceeeccccccceeeeecccccccccc

SEQ       ASPTTDSCDGNASDSSYRTPGIGPVLPLEERGAETETKVQERENGESPLELEQLDQHHEM
SEG       xxxxx.......................................................
PRD       ccccccccccccccccccccccccccccccccchhhhhhhhcccccchhhhhhhhhhhhh

SEQ       KVEHETSS
SEG       ........
PRD       hhhhhccc
```

(No Prosite data available for DKFZphtes3_72k15.3)

(No Pfam data available for DKFZphtes3_72k15.3)

DKFZphtes3_72p16
----

group: intracellular transport and trafficing

DKFZphtes3_72p16 encodes a novel 796 amino acid protein with very strong similarity to Mus musculus maternal-embryonic 3 (Mem3) gene.

Mem3 was isolated from a partial subtraction library of mouse unfertilized eggs and preimplantation embryos. Its transcript is abundant in the unfertilized egg and also actively transcribed from the newly formed zygotic genome. As Mem3, the novel protein is similar to yeast VPS (vacuolar protein sorting) 35. The null allele of VPS35 results in yeast in a differential defect in the sorting of vacuolar carboxypeptidase Y (CPY), proteinase A (PrA), proteinase B (PrB), and alkaline phosphatase (ALP).

The new protein can find application in modulation the sorting of proteins into different compartments.

strong similarity to mouse MEM3 and yeast VPS35

Sequenced by DKFZ

Locus: /map="16p13.3"

Insert length: 2707 bp
Poly A stretch at pos. 2697, no polyadenylation signal found

```
   1 CTACGCGCGG GGCGGGTGCT GCTTGCTGCA GGCTCTGGGG AGTCGCCATG
  51 CCTACAACAC AGCAGTCCCC TCAGGATGAG CAGGAAAAGC TCTTGGATGA
 101 AGCCATACAG GCTGTGAAGG TCCAGTCATT CCAAATGAAG AGATGCCTGG
 151 ACAAAAACAA GCTTATGGAT TCTCTAAAAC ATGCTTCTAA TATGCTTGGT
 201 GAACTCCGGA CTTCTATGTT ATCACCAAAG AGTTACTATG AACTTTATAT
 251 GGCCATTTCT GATGAACTGC ACTACTTGGA GGTCTACCTG ACAGATGAGT
 301 TTGCTAAAGG AAGGAAAGTG GCAGATCTCT ACGAACTTGT ACAGTATGCT
 351 GGAAACATTA TCCCAAGGCT TTACCTTTTG ATCACAGTTG GAGTTGTATA
 401 TGTCAAGTCA TTTCCTCAGT CCAGGAAGGA TATTTTGAAA GATTTGGTAG
 451 AAATGTGCCG TGGTGTGCAA CATCCCTTGA GGGGTCTGTT TCTTCGAAAT
 501 TACCTTCTTC AGTGTACCAG AAATATCTTA CCTGATGAAG GAGAGCCAAC
 551 AGATGAAGAA ACAACTGGTG ACATCAGTGA TTCCATGGAT TTTGTACTGC
 601 TCAACTTTGC AGAAATGAAC AAGCTCTGGG TGCGAATGCA GCATCAGGGA
 651 CATAGCCGAG ATAGAGAAAA AAGAGAACGA GAAAGACAAG AACTGAGAAT
 701 TTTAGTGGGA ACAAATTTGG TGCGCCTCAG TCAGTTGGAA GGTGTAAATG
 751 TGGAACGTTA CAAACAGATT GTTTTGACTG GCATATTGGA GCAAGTTGTA
 801 AACTGTAGGG ATGCTTTGGC TCAAGAATAT CTCATGGAGT GTATTATTCA
 851 GGTTTTCCCT GATGAATTTC ACCTCCAGAC TTTGAATCCT TTTCTTCGGG
 901 CCTGTGCTGA GTTACACCAG AATGTAAATG TGAAGAACAT AATCATTGCT
 951 TTAATTGATA GATTAGCTTT ATTTGCTCAC CGTGAAGATG GACCTGGAAT
1001 CCCAGCGGAT ATTAAACTTT TTGATATATT TTCACAGCAG GTGGCTACAG
1051 TGATACAGTC TAGACAAGAC ATGCCTTCAG AGGATGTTGT ATCTTTACAA
1101 GTCTCTCTGA TTAATCTTGC CATGAAATGT TACCCTGATC GTGTGGACTA
1151 TGTTGATAAA GTTCTAGAAA CAACAGTGGA GATATTCAAT AAGCTCAACC
1201 TTGAACATAT TGCTACCAGT AGTGCAGTTT CAAAGGAACT CACCAGACTT
1251 TTGAAAATAC CAGTTGACAC TTACAACAAT ATTTTAACAG TCTTGAAATT
1301 AAAACATTTT CACCCACTCT TTGAGTACTT TGACTACGAG TCCAGAAAGA
1351 GCATGAGTTG TTATGTGCTT AGTAATGTTC TGGATTATAA CACAGAAATT
1401 GTCTCTCAAG ACCAGGTGGA TTCCATAATG AATTTGGTAT CCACGTTGAT
1451 TCAAGATCAG CCAGATCAAC CTGTAGAAGA CCCTGATCCA GAAGATTTTG
1501 CTGATGAGCA GAGCCTTGTG GGCCGCTTCA TTCATCTGCT GCGCTCTGAG
1551 GACCCTGACC AGCAGTACTT GATTTTGAAC ACAGCACGAA AACATTTTGG
1601 AGCTGGTGGA AATCAGCGGA TTCGCTTCAC ACTGCCACCT TTGGTATTTG
1651 CAGCTTACCA GCTGGCTTTT CGATATAAAG AGAATTCTAA AGTGGATGAC
1701 AAATGGGAAA AGAAATGCCA GAAGATTTTT TCATTTGCCC ACCAGACTAT
1751 CAGTGCTTTG ATCAAAGCAG AGCTGGCAGA ATTGCCCTTA AGACTTTTTC
1801 TTCAAGGAGC ACTAGCTGCT GGGGAAATTG GTTTTGAAAA TCATGAGACA
1851 GTCGCATATG AATTCATGTC CCAGGCATTT TCTCTGTATG AAGATGAAAT
1901 CAGCGATTCC AAAGCACAGC TAGCTGCCAT CACCTTGATC ATTGGCACTT
1951 TTGAAAGGAT GAAGTGCTTC AGTGAAGAGA ATCATGAACC TCTGAGGACT
2001 CAGTGTGCCC TTGCTGCATC CAAACTTCTA AAGAAACCTG ATCAGGGCCG
2051 AGCTGTGAGC ACCTGTGCAC ATCTCTTCTG GTCTGGCAGA AACACGGACA
2101 AAAATGGGGA GGAGCTTCAC GGAGGCAAGA GGGTAATGGA GTGCCTAAAA
2151 AAAGCTCTAA AAATAGCAAA TCAGTGCATG GACCCCTCTC TACAAGTGCA
2201 GCTTTTTATA GAAATTCTGA ACAGATATAT CTATTTTTAT GAAAAGGAAA
2251 ATGATGCGGT AACAATTCAG GTTTTAAACC AGCTTATCCA AAAGATTCGA
2301 GAAGACCTCC CGAATCTTGA ATCCAGTGAA GAAACAGAGC AGATTAACAA
2351 ACATTTTCAT AACACACTGG AGCATTTGCG CTTGCGGCGG GAATCACCAG
2401 AATCCGAGGG GCCAATTTAT GAAGGTCTCA TCCTTTAAAA AGGAAATAGC
2451 TCACCATACT CCTTTCCATG TACATCCAGT GAGGGTTTTA TTACGCTAGG
2501 TTTCCCTTCC ATAGATTGTG CCTTTCAGAA ATGCTGAGGT AGGTTTCCCA
```

```
2551 TTTCTTACCT GTGATGTGTT TTACCCAGCA CCTCCGGACA CTCACCTTCA
2601 GGACCTTAAT AAAATTATTC ACTTGGTAAG TGTTCAAGTC TTTCTGATCA
2651 CCCCAAGTAG CATGACTGAT CTGCAATTTA AAATTCCTGT GATCTGTAAA
2701 AAAAAAA
```

BLAST Results

Entry AC007225 from database EMBLNEW:
Homo sapiens chromosome 16 clone 480G7, WORKING DRAFT SEQUENCE, 38 unordered pieces.
Score = 1081, P = 2.8e-217, identities = 219/221
13 exons

Entry HS015146 from database EMBL:
human STS WI-8848.
Score = 2033, P = 2.9e-87, identities = 425/436

Medline entries

96327632:
Genetic mapping and embryonic expression of a novel, maternally transcribed gene Mem3.

97258867:
Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products.

92360909:
Alternative pathways for the sorting of soluble vacuolar proteins in yeast: a vps35 null mutant missorts and secretes only a subset of vacuolar hydrolases.

10198044:
Distinct Domains within Vps35p Mediate the Retrieval of Two Different Cargo Proteins from the Yeast Prevacuolar/Endosomal Compartment

Peptide information for frame 3

ORF from 48 bp to 2435 bp; peptide length: 796
Category: strong similarity to known protein
Classification: unset

```
  1 MPTTQQSPQD EQEKLLDEAI QAVKVQSFQM KRCLDKNKLM DSLKHASNML
 51 GELRTSMLSP KSYYELYMAI SDELHYLEVY LTDEFAKGRK VADLYELVQY
101 AGNIIPRLYL LITVGVVYVK SFPQSRKDIL KDLVEMCRGV QHPLRGLFLR
151 NYLLQCTRNI LPDEGEPTDE ETTGDISDSM DFVLLNFAEM NKLWVRMQHQ
201 GHSRDREKRE RERQELRILV GTNLVRLSQL EGVNVERYKQ IVLTGILEQV
251 VNCRDALAQE YLMECIIQVF PDEFHLQTLN PFLRACAELH QNVNVKNIII
301 ALIDRLALFA HREDGPGIPA DIKLFDIFSQ QVATVIQSRQ DMPSEDVVSL
351 QVSLINLAMK CYPDRVDYVD KVLETTVEIF NKLNLEHIAT SSAVSKELTR
401 LLKIPVDTYN NILTVLKLKH FHPLFEYFDY ESRKSMSCYV LSNVLDYNTE
451 IVSQDQVDSI MNLVSTLIQD QPDQPVEDPD PEDFADEQSL VGRFIHLLRS
501 EDPDQQYLIL NTARKHFGAG GNQRIRFTLP PLVFAAYQLA FRYKENSKVD
551 DKWEKKCQKI FSFAHQTISA LIKAELAELP LRLFLQGALA AGEIGFENHE
601 TVAYEFMSQA FSLYEDEISD SKAQLAAITL IIGTFERMKC FSEENHEPLR
651 TQCALAASKL LKKPDQGRAV STCAHLFWSG RNTDKNGEEL HGGKRVMECL
701 KKALKIANQC MDPSLQVQLF IEILNRYIYF YEKENDAVTI QVLNQLIQKI
751 REDLPNLESS EETEQINKHF HNTLEHLRLR RESPESEGPI YEGLIL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_72p16, frame 3

TREMBL:AF024504_3 gene: "A_TM017A05.7"; Arabidopsis thaliana BAC TM017A05., N = 2, Score = 927, P = 1.9e-162

PIR:S56936 vacuolar protein-sorting protein VPS35 - yeast (Saccharomyces cerevisiae), N = 3, Score = 826, P = 1.5e-116

TREMBL:MM47024_1 gene: "Mem3"; product: "MEM3"; Mus musculus maternal-embryonic 3 (Mem3) mRNA, complete cds., N = 1, Score = 3376, P = 0

TREMBL:S42186_1 gene: "VPS35"; product: "Vps35p"; VPS35=vacuolar protein sorting [Saccharomyces cerevisiae=yeast, Genomic, 3790 nt], N = 3, Score = 813, P = 4.4e-115

```
>TREMBL:MM47024_1 gene: "Mem3"; product: "MEM3"; Mus musculus
     maternal-embryonic 3 (Mem3) mRNA, complete cds.
            Length = 754

  HSPs:

 Score = 3376 (506.5 bits), Expect = 0.0e+00, P = 0.0e+00
 Identities = 666/721 (92%), Positives = 682/721 (94%)

Query:    78 EVYLTDEFAKGRKVADLYELVQYAGNIIPRLYLLITVGVVYVKSFPQSRKDILKDLVEMC 137
             +VYLTDEFAKG ++ADLYELVQY+GNIIPRLYLLITVGVVYVKSFPQSRKDILKDLVEMC
Sbjct:    34 KVYLTDEFAKGERLADLYELVQYSGNIIPRLYLLITVGVVYVKSFPQSRKDILKDLVEMC 93

Query:   138 RGVQHPLRGLFLRNYLLQCTRNILPDEGEPTDEETTGDISDSMDFVLLNFAEMNKLWVRM 197
             RGVQHPLRGLFLRNYLLQCTRNILPDEGEPTDEETTGDISDSMDFVLLNFAEMNKLWVRM
Sbjct:    94 RGVQHPLRGLFLRNYLLQCTRNILPDEGEPTDEETTGDISDSMDFVLLNFAEMNKLWVRM 153

Query:   198 QHQGHSRDREKRERERQELRILVGTNLVRLSQLEG-VNVERYKQIVLTGILEQVVNCRDA 256
             QHQGHSRDREKRERERQELRILVGTNLV L+ +         +QIVLTGILEQVVNCRDA
Sbjct:   154 QHQGHSRDREKRERERQELRILVGTNLVALTLVSWRCKCGTLQQIVLTGILEQVVNCRDA 213

Query:   257 LAQEYLMECIIQVFPDEFHLQTLNPFLRACAELHQNVNVKNIIIALIDRLALFAHREDGP 316
             LAQE  MECIIQVFPDEFHLQTLNPFLRACAELHQNVNVKNIIIALIDRLALFAHRE  P
Sbjct:   214 LAQEISMECIIQVFPDEFHLQTLNPFLRACAELHQNVNVKNIIIALIDRLALFAHREMEP 273

Query:   317 GIPADIKLFDIFSQQVATVIQSRQDMPSEDVVSLQVSLINLAMKCYPDRVDYVDKVLETT 376
             GIPA++KLFDIFSQQVATVIQSR+DMPSEDVVSLQVSLINLAMKCYPDRVDYVDKVLETT
Sbjct:   274 GIPAELKLFDIFSQQVATVIQSRRDMPSEDVVSLQVSLINLAMKCYPDRVDYVDKVLETT 333

Query:   377 VEIFNKLNLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHFHPLFEYFDYESR--K 434
             VEIFNKLNLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHFHPLFEYFDYES   K
Sbjct:   334 VEIFNKLNLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHFHPLFEYFDYESSPGK 393

Query:   435 SMSCYVLSNVLDYNTEIVSQDQVDSIMNLVSTLIQDQPDQPVEDPDPEDFADEQSLVGRF 494
             SMSCYVLSNVLDYNTEIVSQDQVDSIMNLVSTLIQDQPDQPVEDPDPEDFADEQSLVGRF
Sbjct:   394 SMSCYVLSNVLDYNTEIVSQDQVDSIMNLVSTLIQDQPDQPVEDPDPEDFADEQSLVGRF 453

Query:   495 IHLLRSEDPDQQYLILNTARKHFGAGGNQRIRFTLPPLVFAAYQLAFRYKENSKVDDKWE 554
             IHLLRS+DPDQQYLILNTARKHFGAGGNQRIRFTLPPLVFAAYQLAFRYKENSK     +
Sbjct:   454 IHLLRSDDPDQQYLILNTARKHFGAGGNQRIRFTLPPLVFAAYQLAFRYKENSKWMTSGK 513

Query:   555 KKCQKIFSFAHQTISALIKAELAELPLRLFLQGALAAGEIGFENHETVAYEFMSQAFSLY 614
             +  ++ F   HQTISALIKAELAELPLRLFLQGALAAGEIGFENHETVAYEFMSQAFSLY
Sbjct:   514 RNARRYFHLPHQTISALIKAELAELPLRLFLQGALAAGEIGFENHETVAYEFMSQAFSLY 573

Query:   615 EDEISDSKAQLAAITLIIGTFERMKCFSEENHEPLRTQCALAASKLLKKPDQGRAVSTCA 674
             EDEISDSKAQLAAITLIIGTFERMKCFSEENHEPLRT+CALAASKLLKKPDQ      C
Sbjct:   574 EDEISDSKAQLAAITLIIGTFERMKCFSEENHEPLRTECALAASKLLKKPDQAEREHMCT 633

Query:   675 HLFWSGRNTDKNGEELHGGKRVMECLKKALKIANQCMDPSLQVQLFIEILNRYIYFYEKE 734
              L WSGRNTDKNGEELHGGKRVMECLKKALKIANQCMDPSLQVQLFIEILNRYIYFYEKE
Sbjct:   634 SL-WSGRNTDKNGEELHGGKRVMECLKKALKIANQCMDPSLQVQLFIEILNRYIYFYEKE 692

Query:   735 NDAVTIQVLNQLIQKIREDLPNLESSEETEQINKHFHNTLEHLRLRRESPESEGPIYEGL 794
             NDAVTIQVLNQLIQKIREDLPNLESSEETEQINKHFHNTLEHLR RRESPESEGPIYEGL
Sbjct:   693 NDAVTIQVLNQLIQKIREDLPNLESSEETEQINKHFHNTLEHLRTRRESPESEGPIYEGL 752

Query:   795 IL 796
             IL
Sbjct:   753 IL 754
```

Pedant information for DKFZphtes3_72p16, frame 3

Report for DKFZphtes3_72p16.3

[LENGTH] 796

[MW] 91723.67
[pI] 5.32
[HOMOL] TREMBL:MM47024_1 gene: "Mem3"; product: "MEM3"; Mus musculus maternal-embryonic 3 (Mem3) mRNA, complete cds. 0.0
[FUNCAT] 30.25 vacuolar and lysosomal organization [S. cerevisiae, YJL154c] 1e-110
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YJL154c] 1e-110
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YJL154c] 1e-110
[FUNCAT] 30.22 endosomal organization [S. cerevisiae, YJL154c] 1e-110
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YJL154c] 1e-110
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YJL154c] 1e-110
[FUNCAT] 09.07 biogenesis of endoplasmatic reticulum [S. cerevisiae, YJL154c] 1e-110
[BLOCKS] BL01092Q
[PIRKW] yeast vacuole 1e-108
[PIRKW] membrane protein 1e-108
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 5.40 %

```
SEQ  MPTTQQSPQDEQEKLLDEAIQAVKVQSFQMKRCLDKNKLMDSLKHASNMLGELRTSMLSP
SEG  ............................................................
PRD  cccccccccchhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhhcccc
MEM  ............................................................

SEQ  KSYYELYMAISDELHYLEVYLTDEFAKGRKVADLYELVQYAGNIIPRLYLLITVGVVYVK
SEG  ............................................................
PRD  cceeeeehhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhcccccccceeeeeceeeee
MEM  ..............................................MMMMMMMMMMMMMM

SEQ  SFPQSRKDILKDLVEMCRGVQHPLRGLFLRNYLLQCTRNILPDEGEPTDEETTGDISDSM
SEG  .........................................xxxxxxxxxxxxxx.....
PRD  ecccchhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhccccccccccccccccccccch
MEM  MMMMMMMMMM..................................................

SEQ  DFVLLNFAEMNKLWVRMQHQGHSRDREKREREREQELRILVGTNLVRLSQLEGVNVERYKQ
SEG  .......................xxxxxxxxxxxxxx.......................
PRD  hhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhcchhhhhhhcchhhhhhhh
MEM  ............................................................

SEQ  IVLTGILEQVVNCRDALAQEYLMECIIQVFPDEFHLQTLNPFLRACAELHQNVNVKNIII
SEG  ............................................................
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhcccchhhhhhh
MEM  ............................................................

SEQ  ALIDRLALFAHREDGPGIPADIKLFDIFSQQVATVIQSRQDMPSEDVVSLQVSLINLAMK
SEG  ............................................................
PRD  hhhhhhhhhhhcccccccccchhhhhhhhhhhhhhhhhcccccccchhhhhhhhhhhhhh
MEM  ............................................................

SEQ  CYPDRVDYVDKVLETTVEIFNKLNLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKH
SEG  ............................................................
PRD  ccccccchhhhhhhhhhhhhhhcchhhhhhhcchhhhhhhhhcccccchhhhhhhhhhhhh
MEM  ............................................................

SEQ  FHPLFEYFDYESRKSMSCYVLSNVLDYNTEIVSQDQVDSIMNLVSTLIQDQPDQPVEDPD
SEG  ................................................xxxxxxxxxxxx
PRD  hhhheeecccchhhhhhhhhhhccccceeehhhhhhhhhhhhhhhhhccccccccccccc
MEM  ............................................................

SEQ  PEDFADEQSLVGRFIHLLRSEDPDQQYLILNTARKHFGAGGNQRIRFTLPPLVFAAYQLA
SEG  xxx.........................................................
PRD  cccchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhcccccceeeeeccchhhhhhhhh
MEM  ............................................................

SEQ  FRYKENSKVDDKWEKKCQKIFSFAHQTISALIKAELAELPLRLFLQGALAAGEIGFENHE
SEG  ............................................................
PRD  hhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
MEM  ............................................................

SEQ  TVAYEFMSQAFSLYEDEISDSKAQLAAITLIIGTFERMKCFSEENHEPLRTQCALAASKL
SEG  ............................................................
PRD  eeeehhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhcccccccchhhhhhhhhhhhhh
MEM  ............................................................

SEQ  LKKPDQGRAVSTCAHLFWSGRNTDKNGEELHGGKRVMECLKKALKIANQCMDPSLQVQLF
SEG  ............................................................
PRD  hhcccceeeeecccccccccccccccccccccchhhhhhhhhhhhhhhhhhhchhhhhhhh
MEM  ............................................................

SEQ  IEILNRYIYFYEKENDAVTIQVLNQLIQKIREDLPNLESSEETEQINKHFHNTLEHLRLR
```

```
SEG     ............................................................
PRD     hhhhhhhhhhhccccceeeeehhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhh
MEM     ............................................................

SEQ     RESPESEGPIYEGLIL
SEG     ................
PRD     hhcccccccceeeccc
MEM     ................
```

(No Prosite data available for DKFZphtes3_72p16.3)

(No Pfam data available for DKFZphtes3_72p16.3)

DKFZphtes3_7b22
----

group: cell structure and motility

DKFZphtes3_7b22 encodes a novel 443 amino acid protein with weak similarity to paramyosins.

The novel protein is related to paramyosin, a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as Schistosoma mansoni.

The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamic.

similarity to paramyosins

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: /map="3"

Insert length: 2291 bp
Poly A stretch at pos. 2241, polyadenylation signal at pos. 2213

```
   1 GGAAGAAAGG CTAGCGGGCG TTGGCCGTAT GTGGGTGTCT TGAGGCAGTT
  51 TTTCAGTTCT TTCATTTACC AAAGTGACAT GCACCTACTA GGTGCCAGGT
 101 GTTTAGACGT ACATACAACC CTCTGCAAAA TCTTTCAGTG TAGTCCTCTG
 151 TATGAAAAGT TTCCAGCCAA GAATTGCCAC TGCACCTGAG ATAAGGGGGA
 201 TCCTGGCCAT TAAGGAAACC TTGCCTTCGA AACTGAGCCG TGAGGAACTA
 251 TACAAAATGG GAAATTGGGA CAAATCCCAG TGGCTCATGA CACTAAGAAG
 301 TAAAATTACG AACTCACTGA GCTGGAAGTC ATTCAACGGG AATTGAATAG
 351 GTAACTGCAC TTTTGTGAGA TTATAAATAT ACCACGGAGG GTAACGAAGC
 401 TACAGAAGAA TGGAAGAAGA CAGCCTGGAA GACTCAAACC TTCCTCCAAA
 451 AGTTTGGCAT TCTGAGATGA CGGTGTCAGT GACAGGCGAA CCACCTAGTA
 501 CCGTAGAAGA AGAAGGAATA CCTAAAGAAA CAGACATAGA AATCATCCCA
 551 GAAATCCCGG AAACTCTAGA GCCACTGTCC CTTCCAGATG TGCTGAGGAT
 601 CTCGGCAGTT CTGGAGGACA CCACAGACCA GCTCTCTATT CTGAACTACA
 651 TCATGCCCGT TCAGTACGAA GGGAGACAGA GCATCTGCGT GAAAAGCAGA
 701 GAAATGAATC TAGAAGGAAC GAATCTAGAC AAACTTCCAA TGGCCTCAAC
 751 AATCACAAAA ATACCCAGTC CGTTAATAAC TGAGGAAGGA CCCAACTTGC
 801 CAGAAATCAG ACACAGAGGC CGGTTCGCTG TGGAGTTTAA CAAAATGCAG
 851 GATCTTGTCT TCAAAAAACC TACAAGGCAG ACCATCATGA CTACGGAGAC
 901 ACTGAAGAAA ATTCAGATTG ATAGGCAGTT TTTCAGCGAT GTGATTGCAG
 951 ATACCATTAA GGAGTTGCAA GATTCGGCCA CTTACAACAG TCTCCTGCAA
1001 GCTTTGAGCA AAGAGAGGGA AAACAAAATG CATTTCTATG ACATCATTGC
1051 CAGGGAGGAA AAAGGAAGAA AACAGATAAT ATCACTTCAA AAACAGCTAA
1101 TTAATGTCAA AAAGGAATGG CAATTTGAAG TCCAGAGTCA GAATGAGTAT
1151 ATTGCTAACC TCAAGGACCA ACTGCAAGAG ATGAAGGCAA AATCCAACTT
1201 GGAGAATCGC TACATGAAAA CCAATACCGA GCTGCAGATT GCCCAGACCC
1251 AGAAAAAGTG TAACAGAACA GAGGAACTCT TGGTGGAAGA GATTGAGAAA
1301 CTCAGGATGA AAACCGAAGA AGAGGCCCGG ACTCATACAG AGATTGAAAT
1351 GTTCCTTAGA AAGGAGCAGC AGAAACTTGA GGAGAGGCTG GAGTTCTGGA
1401 TGGAGAAATA CGATAAGGAC ACAGAAATGA AACAGAATGA ACTAAATGCT
1451 CTCAAAGCCA CAAAGGCCAG TGACTTAGCA CACCTTCAAG ACCTGGCAAA
1501 GATGATAAGA GAGTATGAAC AGGTCATCAT TGAAGATCGT ATAGAAAAGG
1551 AGAGGAGCAA GAAGAAGGTA AAACAGGATC TCTTGGAATT AAAGAGCGTT
1601 ATAAAGCTCC AGGCCTGGTG GCGAGGCACT ATGATACGGA GAGAAATTGG
1651 TGGTTTCAAG ATGCCTAAAG ACAAAGTTGA TAGCAAGGAT TCAAAAGGCA
1701 AAGGTAAAGG CAAGGATAAG AGGAGAGGCA AGAAGAAGTG ACCAAGTTCT
1751 CTTTTGTGTT TTCTGCTGGT ATTCTGGAGG TGGGAAGGAC TTGGAGAGTT
1801 AAGAAACACC TGGTACCTCA AAGATGACTC ATCTACAGGT TGTTTCCTAT
1851 TGAGACTTTC CCAGGGAAGC CTGATTTCAC TTTGCCTGTT AATTTCACTC
1901 TGCCTGTTAG GTGGGTTTTC AAACCCTGAT TTAGGATTAC ACCATTGACT
1951 TAGGGCTTCC TCATACCTTG CTGGGAAGAA GTTTCTAGTA GTCCTGTGAA
2001 GATTCATTCT TCTTGCTCTT TCTCAGCAGA ACAAAGGAGT TCACTGGCTT
2051 AGCTACAGTG ACGCATTGAA ACTTGAGTAA TTCCTGTAAT GTCAGATTTT
2101 GATTTTACCC AATTTGTCTG TAGTGAAAAA ACTCTTATGA GCAAAAGTAT
2151 TCAGTAGGAA TTACAATATG ATGTTATTAG CTGTCCAGCA TAATATATAC
2201 ACAGCAAAGT TTTAATAAAT GTTGGTTCCT GCCTGCCTTT TAAAAAAAAA
2251 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA A
```

BLAST Results

Entry G36731 from database EMBL:
SHGC-52923 Human Homo sapiens STS cDNA.

Score = 2262, P = 1.3e-97, identities = 462/468

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 410 bp to 1738 bp; peptide length: 443
Category: similarity to known protein

```
  1 MEEDSLEDSN LPPKVWHSEM TVSVTGEPPS TVEEEGIPKE TDIEIIPEIP
 51 ETLEPLSLPD VLRISAVLED TTDQLSILNY IMPVQYEGRQ SICVKSREMN
101 LEGTNLDKLP MASTITKIPS PLITEEGPNL PEIRHRGRFA VEFNKMQDLV
151 FKKPTRQTIM TTETLKKIQI DRQFFSDVIA DTIKELQDSA TYNSLLQALS
201 KERENKMHFY DIIAREEKGR KQIISLQKQL INVKKEWQFE VQSQNEYIAN
251 LKDQLQEMKA KSNLENRYMK TNTELQIAQT QKKCNRTEEL LVEEIEKLRM
301 KTEEEARTHT EIEMFLRKEQ QKLEERLEFW MEKYDKDTEM KQNELNALKA
351 TKASDLAHLQ DLAKMIREYE QVIIEDRIEK ERSKKKVKQD LLELKSVIKL
401 QAWWRGTMIR REIGGFKMPK DKVDSKDSKG KGKGKDKRRG KKK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7b22, frame 2

SWISSPROT:MYSP_BRUMA PARAMYOSIN., N = 1, Score = 158, P = 5.8e-08

PIR:A44972 paramyosin - nematode (Dirofilaria immitis) (fragment), N = 1, Score = 157, P = 7.1e-08

SWISSPROT:MYSP_ONCVO PARAMYOSIN., N = 1, Score = 157, P = 7.4e-08

PIR:S52537 emm L 15 protein - Streptococcus pyogenes, N = 1, Score = 151, P = 8.6e-08

>SWISSPROT:MYSP_BRUMA PARAMYOSIN.
Length = 880

HSPs:

Score = 158 (23.7 bits), Expect = 5.8e-08, P = 5.8e-08
Identities = 66/259 (25%), Positives = 125/259 (48%)

```
Query:   142 EFNKMQDLVFKKPTRQTIMTTETLKKIQIDRQFFSDVIADTIKELQDSATYNSLLQALSK 201
             + K + L  K   R    T E   K++   +  +D +A   + LQ  A  N LL+ +
Sbjct:   169 QLKKDKHLAEKAAERFEAQTVELSNKVEDLNRHVND-LAQQRQRLQ--AENNDLLKEIHD 225

Query:   202 ER---ENKMHF-YDIIAREEKGRKQIISLQKQLINVKKEWQFEVQSQNEYIANLKDQLQE 257
             ++    +N  H  Y +  + E+ R+++   +++   ++ +    +VQ + + +    D+  E
Sbjct:   226 QKVQLDNLQHVKYQLAQQLEEARRRLEDAERERSQLQAQLH-QVQLELDSVRTALDE--E 282

Query:   258 MKAKSNLENRYMKTNTELQIAQTQKKCNRTEELLVEEIEKLRMKT-EEEARTHTEIEMFL 316
               A++  E++     NTE  I Q + K +     L  EE+E LR K  +++A     +IE+ L
Sbjct:   283 SAARAEAEHKLALANTE--ITQWKSKFDAEVALHHEEVEDLRKKMLQKQAEYEEQIEIML 340

Query:   317 RKEQQ--KLEERLEFWMEKYDKDTEMKQNELNALKATKASDLAHLQDLAKMIREYEQVII 374
             +K  Q  K + RL+  +E     D E  QN +  L+  K        + L K + E +  I
Sbjct:   341 QKISQLEKAKSRLQSEVEVLIVDLEKAQNTIAILERAK-------EQLEKTVNELKVRID 393

Query:   375 EDRIEKERSKKKVKQDLLELKSVIKL 400
             E  +E E ++++ +  L EL+ +  L
Sbjct:   394 ELTVELEAAQREARAALAELQKLKNL 419
```

Score = 118 (17.7 bits), Expect = 1.3e-03, P = 1.3e-03
Identities = 54/231 (23%), Positives = 108/231 (46%)

```
Query:   181 DTIKELQDSATYNSLLQ----ALSKERENKMHFYDIIAREEKG-RKQIISLQKQLINVKK 235
             D +KE+ D        LQ     L+++ E       +    RE        + Q+  +Q +L +V+
Sbjct:   218 DLLKEIHDQKVQLDNLQHVKYQLAQQLEEARRRLEDAERERSQLQAQLHQVQLELDSVRT 277
```

```
Query:   236 EWQFE--VQSQNEY-IANLKDQLQEMKAKSNLENRYMKTNTE-LQIAQTQKKCNRTEELL 291
             E    +++ E+ +A    ++ + K+K + E          E L+     QK+     E++
Sbjct:   278 ALDEESAARAEAEHKLALANTEITQWKSKFDAEVALHHEEVEDLRKKMLQKQAEYEEQIE 337

Query:   292 VEEIEKLRMKTEEEARTHTEIEMF---LRKEQQKLE--ERLEFWMEKYDKDTEMKQNELN 346
             +  ++K+     + ++R  +E+E+      L K Q  +   ER +  +EK   + +++ +EL
Sbjct:   338 IM-LQKISQLEKAKSRLQSEVEVLIVDLEKAQNTIAILERAKEQLEKTVNELKVRIDELT 396

Query:   347 A-LKATKASDLAHLQDLAKMIREYEQVIIEDRIEKERSKKKVKQDLLELKSVI 398
              L+A +     A L +L K+   YE+ + E +     R  KK++ DL E K  +
Sbjct:   397 VELEAAQREARAALAELQKLKNLYEKAV-EQKEALARENKKLQDDLHEAKEAL 448

 Score = 107 (16.1 bits), Expect = 2.1e-02, P = 2.1e-02
 Identities = 49/279 (17%), Positives = 124/279 (44%)

Query:   123 ITEEGPNLPEIRHRGRFAV-EFNKMQDLVFKKPTRQTIMTTETLKKIQIDRQFFSDVIAD 181
             I E     L    +    R A+ E  K+++L  K    ++   +  E  KK+Q D     + +AD
Sbjct:   392 IDELTVELEAAQREARAALAELQKLKNLYEKAVEQKEALAREN-KKLQDDLHEAKEALAD 450

Query:   182 TIKELQDSATYNSLLQALSKERENKMHFYDIIAREEKGRKQ--IISLQKQLINVKKEWQF 239
               ++L +     N+ L     +E +  +    +    R+ + R Q  +  LQ+  I  +++  Q
Sbjct:   451 ANRKLHELDLENARLAGEIRELQTALKESEAARRDAENRAQRALAELQQLRIEMERRLQE 510

Query:   240 EVQSQNEYIANLKDQLQEMKAKSNLENRYMKTNTELQIAQTQKKCNRTE-ELLVEEIEKL 298
             + +          N++ ++ + A    L +   + E+      + + +    E E+ V+ + +
Sbjct:   511 KEEEMEALRKNMQFEIDRLTAA--LADAEARMKAEISRLKKKYQAEIAELEMTVDNLNRA 568

Query:   299 RMKTEEEARTHTEIEMFLRKEQQKLEERLEFWMEKYDKDTEMKQNELNALKATKASDLAH 358
             ++ ++  +  +E    L+   +   + +L+  +++Y        + Q +++AL A +   +
Sbjct:   569 NIEAQKTIKKQSEQLKILQASLEDTQRQLQQTLDQY----ALAQRKVSALSA-ELEECKV 623

Query:   359 LQDLAKMIREYEQVIIEDRIEKERSKKKVKQDLLELKSVIKLQ 401
             D A    R+  ++ +E+     +      V   +L  +K+ ++ +
Sbjct:   624 ALDNAIRARKQAEIDLEEANGRITDLVSVNNNLTAIKNKLETE 666
```

Pedant information for DKFZphtes3_7b22, frame 2

Report for DKFZphtes3_7b22.2

```
[LENGTH]       443
[MW]           51917.95
[pI]           6.18
[HOMOL]        PIR:S28589 trichohyalin - rabbit 2e-08
[FUNCAT]       30.03 organization of cytoplasm     [S. cerevisiae, YDL058w] 7e-07
[FUNCAT]       08.07 vesicular transport (golgi network, etc.)     [S. cerevisiae, YDL058w]
7e-07
[FUNCAT]       1 genome replication, transcription, recombination and repair     [M.
jannaschii, MJ1322] 5e-06
[FUNCAT]       03.22 cell cycle control and mitosis  [S. cerevisiae, YPR141c] 1e-05
[FUNCAT]       03.13 meiosis  [S. cerevisiae, YPR141c] 1e-05
[FUNCAT]       11.01 stress response  [S. cerevisiae, YPR141c] 1e-05
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
      [S. cerevisiae, YPR141c] 1e-05
[FUNCAT]       08.22 cytoskeleton-dependent transport     [S. cerevisiae, YPR141c] 1e-05
[FUNCAT]       09.10 nuclear biogenesis     [S. cerevisiae, YPR141c] 1e-05
[FUNCAT]       30.05 organization of centrosome     [S. cerevisiae, YPR141c] 1e-05
[FUNCAT]       06.10 assembly of protein complexes  [S. cerevisiae, YPR141c] 1e-05
[FUNCAT]       99 unclassified proteins     [S. cerevisiae, YOR216c] 3e-05
[FUNCAT]       11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair)        [S. cerevisiae, YKR095w] 6e-05
[FUNCAT]       30.10 nuclear organization     [S. cerevisiae, YKR095w] 6e-05
[FUNCAT]       30.02 organization of plasma membrane     [S. cerevisiae, YER008c] 1e-04
[FUNCAT]       08.16 extracellular transport     [S. cerevisiae, YER008c] 1e-04
[FUNCAT]       03.04 budding, cell polarity and filament formation  [S. cerevisiae, YER008c]
1e-04
[FUNCAT]       30.04 organization of cytoskeleton   [S. cerevisiae, YDR356w] 2e-04
[FUNCAT]       08.01 nuclear transport     [S. cerevisiae, YDL207w] 4e-04
[FUNCAT]       04.07 rna transport    [S. cerevisiae, YDL207w] 4e-04
[FUNCAT]       06.07 protein modification (glycolsylation, acylation, myristylation,
palmitylation, farnesylation and processing)     [S. cerevisiae, YKL201c] 5e-04
[EC]           3.6.1.32 Myosin ATPase 3e-08
[PIRKW]        phosphotransferase 6e-06
[PIRKW]        citrulline 8e-06
[PIRKW]        tandem repeat 1e-07
[PIRKW]        heart 6e-06
[PIRKW]        polymorphism 4e-06
[PIRKW]        serine/threonine-specific protein kinase 6e-06
[PIRKW]        DNA binding 8e-08
```

```
[PIRKW]        muscle contraction 1e-07
[PIRKW]        actin binding 3e-08
[PIRKW]        ATP 3e-08
[PIRKW]        thick filament 1e-07
[PIRKW]        phosphoprotein 3e-08
[PIRKW]        glycoprotein 4e-06
[PIRKW]        skeletal muscle 1e-07
[PIRKW]        calcium binding 8e-06
[PIRKW]        alternative splicing 3e-08
[PIRKW]        coiled coil 3e-08
[PIRKW]        P-loop 3e-08
[PIRKW]        heptad repeat 4e-06
[PIRKW]        methylated amino acid 3e-08
[PIRKW]        basement membrane 4e-06
[PIRKW]        cardiac muscle 6e-06
[PIRKW]        extracellular matrix 4e-06
[PIRKW]        hydrolase 3e-08
[PIRKW]        membrane protein 4e-06
[PIRKW]        EF hand 8e-06
[PIRKW]        cytoskeleton 8e-06
[PIRKW]        hair 8e-06
[SUPFAM]       myosin heavy chain 3e-08
[SUPFAM]       unassigned Ser/Thr or Tyr-specific protein kinases 6e-06
[SUPFAM]       calmodulin repeat homology 8e-06
[SUPFAM]       myosin motor domain homology 3e-08
[SUPFAM]       trichohyalin 8e-06
[SUPFAM]       protein kinase homology 6e-06
[PROSITE]      AMIDATION        2
[PROSITE]      CAMP_PHOSPHO_SITE        1
[PROSITE]      CK2_PHOSPHO_SITE         12
[PROSITE]      TYR_PHOSPHO_SITE         2
[PROSITE]      PKC_PHOSPHO_SITE         4
[PROSITE]      ASN_GLYCOSYLATION        1
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY     10.61 %

SEQ   MEEDSLEDSNLPPKVWHSEMTVSVTGEPPSTVEEEGIPKETDIEIIPEIPETLEPLSLPD
SEG   ....................................xxxxxxxxxxxxxxxxxxxxxxx.
PRD   cccccccccccccccccceeeeecccccccceeeeecccccceeeeeeecccccccccccc

SEQ   VLRISAVLEDTTDQLSILNYIMPVQYEGRQSICVKSREMNLEGTNLDKLPMASTITKIPS
SEG   ............................................................
PRD   chhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   PLITEEGPNLPEIRHRGRFAVEFNKMQDLVFKKPTRQTIMTTETLKKIQIDRQFFSDVIA
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   DTIKELQDSATYNSLLQALSKERENKMHFYDIIAREEKGRKQIISLQKQLINVKKEWQFE
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   VQSQNEYIANLKDQLQEMKAKSNLENRYMKTNTELQIAQTQKKCNRTEELLVEEIEKLRM
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   KTEEEARTHTEIEMFLRKEQQKLEERLEFWMEKYDKDTEMKQNELNALKATKASDLAHLQ
SEG   ............................................................
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ   DLAKMIREYEQVIIEDRIEKERSKKKVKQDLLELKSVIKLQAWWRGTMIRREIGGFKMPK
SEG   ...........................................................x
PRD   hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccc

SEQ   DKVDSKDSKGKGKGKDKRRGKKK
SEG   xxxxxxxxxxxxxxxxxxxxxxx
PRD   ccccccccccccccccccccccc
```

Prosite for DKFZphtes3_7b22.2

PS00001	285->289	ASN_GLYCOSYLATION	PDOC00001
PS00004	152->156	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00005	182->185	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	383->386	PKC_PHOSPHO_SITE	PDOC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDOC00006
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006

PS00006	41->45	CK2_PHOSPHO_SITE	PDOC00006
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	182->186	CK2_PHOSPHO_SITE	PDOC00006
PS00006	243->247	CK2_PHOSPHO_SITE	PDOC00006
PS00006	262->266	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	302->306	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	310->314	CK2_PHOSPHO_SITE	PDOC00006
PS00007	261->269	TYR_PHOSPHO_SITE	PDOC00007
PS00007	184->193	TYR_PHOSPHO_SITE	PDOC00007
PS00009	218->222	AMIDATION	PDOC00009
PS00009	439->443	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3_7b22.2)

DKFZphtes3_7d17
----

group: testes derived

DKFZphtes3_7d17 encodes a novel 633 amino acid protein with weak similarity to human KIAA0454.

Pfam predicts a TNFR/NGFR cysteine-rich region.
No informative BLAST results; No predictive prosite or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to KIAA0454

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3608 bp
Poly A stretch at pos. 3587, polyadenylation signal at pos. 3570

```
   1 GGGAAGTTAC GGCGAAGTCC ACCCAGCGTT TCTCAGGCAA TCTGAAGGCA
  51 AATCCTGTTT AGACCCAGGC GAAGGTTCCT GGTGACCCAG GCTCTCACCA
 101 GCCAATTGTC CCTTGCCGTC CTCCTGAGGG TATCTGGAGC TTCAGTGCTG
 151 TGTGCTCTTG GCCTCCACAC TGGGGATGCC ACTGACTCCC ACTGTCCAGG
 201 GCTTCCAGTG GACTCTCCGA GGCCCTGATG TAGAAACTTC CCCATTCGGT
 251 GCACCAAGAG CAGCCTCACA TGGTGTGGGC CGACATCAAG AGCTGCGAGA
 301 TCCAACAGTC CCTGGCCCCA CCTCTTCTGC CACAAACGTC AGCATGGTGG
 351 TATCTGCCGG CCCTTGGTCC GGTGAGAAGG CAGAGATGAA CATTCTAGAA
 401 ATCAACAAGA AATCGCGCCC CCAGCTGGCA GAGAACAAAC AGCAGTTCAG
 451 AAACCTCAAA CAGAAATGTC TTGTAACTCA AGTGGCCTAC TTCCTGGCCA
 501 ACCGGCAAAA TAATTACGAC TATGAAGACT GCAAAGACCT CATAAAATCT
 551 ATGCTGAGGG ATGAGCGGCT GCTCACAGAA GAGAAGCTTG CAGAGGAGCT
 601 CGGGCAAGCT GAGGAGCTCA GGCAATATAA AGTCCTGGTT CACTCTCAGG
 651 AACGAGAGCT GACCCAGTTA AGGGAGAAGT TACAGGAAGG GAGAGATGCC
 701 TCCCGCTCAT TGAATCAGCA TCTCCAGGCC CTCCTCACTC CGGATGAGCC
 751 GGACAACTCC CAGGGACGGG ACCTCCGAGA ACAGCTGGCT GAGGGATGTA
 801 GGCTGGCACA GCACCTCGTC CAAAAGCTCA GCCCAGAAAA TGATGACGAT
 851 GAGGATGAAG ATGTTAAAGT TGAGGAGGCT GAGAAAGTAC AGGAATTATA
 901 TGCCCCCAGG GAGGTGCAGA AGGCTGAAGA AAAGGAAGTC CCTGAGGACT
 951 CACTGGAGGA GTGTGCCATC ACTTGTTCAA ATAGCCACCA CCCTTGTGAG
1001 TCCAACCAGC CTTACGGGAA CACCAGAATC ACATTTGAGG AAGACCAAGT
1051 CGACTCAACT CTCATTGACT CATCCTCTCA TGATGAATGG TTGGATGCTG
1101 TATGCATTAT CCCAGAAAAT GAAAGTGATC ATGAGCAAGA GGAAGAAAAA
1151 GGGCCAGTGT CTCCCAGGAA TCTGCAGGAG TCTGAAGAGG AGGAAGCCCC
1201 CCAGGAGTCC TGGGATGAAG GTGATTGGAC TCTCTCAATT CCTCCTGACA
1251 TGTCTGCCTC ATACCAGTCT GACAGGAGCA CCTTTCACTC AGTAGAGGAA
1301 CAGCAAGTCG GCTTGGCTCT TGACATAGGC AGACATTGGT GTGATCAAGT
1351 GAAAAAGGAG GACCAAGAGG CCACAAGTCC CAGGCTCAGC AGGGAGCTGC
1401 TGGATGAGAA AGAGCCTGAA GTCTTGCAGG ACTCACTGGA TAGATTTTAT
1451 TCAACTCCTT TTGAGTACCT GGAACTGCCT GACTTATGCC AGCCCTACAG
1501 AAGTGACTTT TACTCATTGC AGGAACAACA CCTTGGCTTG GCTCTTGACT
1551 TGGACAGAAT GAAAAAGGAC CAAGAAGAGG AAGAAGACCA AGGCCCACCA
1601 TGCCCCAGGC TCAGCAGAGA GCTGCCGGAG GTAGTAGAGC CTGAGGACTT
1651 GCAGGACTCA CTGGATAGAT GGTATTCGAC TCCTTTCAGT TATCCAGAAC
1701 TGCCTGATTC ATGCCAGCCC TACGGAAGTT GCTTTTACTC ATTGGAGGAA
1751 GAACACGTTG GCTTTTCTCT TGACGTGGAT GAAATTGAAA AGTACCAAGA
1801 AGGGGAAGAA GATCAAAAGC CACCATGCCC CAGGCTCAAC GAGGTGCTGA
1851 TGGAAGCAGA AGAGCCTGAA GTCTTGCAGG ACTCACTGGA TAGATGTTAT
1901 TCGACTACTT CAACTTACTT TCAACTACAT GCCTCATTCC AGCAGTACAG
1951 AAGTGCCTTT TACTCATTTG AGGAACAGGA CGTCAGCTTG GCCCTTGACG
2001 TGGACAATAG GTTTTTTACT TTGACAGTGA TAAGGCACCA CCTGGCCTTC
2051 CAGATGGGAG TCATATTCCC ACACTAAGCA GCCCTTACTA AGCTGAGAGA
2101 TGTCATTGCT GCAGGCAGGA CCTATAGGCA CATGTAGGTT TGAATGAAAC
2151 TGTAGTTCCC TTTGGAAGCC CAGTCATAGG ATGGGAAAGT GGGCATGGCT
2201 CTATTCCTAT TCTCAGACCA TGCCAGTGGC CACCTGTGCT CAGTCTGAAG
2251 ACGTTGGACC CAAGTTAGGT GTGACACGTT CACACGACTA TGTAGCACAT
2301 GCCGGGAGTG ATCTGCCAGA CATTCTAATT TGAACCAGAT ATCTCTGGGT
2351 AGCTACAAAG TTCCTCAGGG GTTTCATTTT GCAGGCATGT CTCTGAGCTT
2401 CTATACCTGC TCAAGGTCAG TGTCATCTTT GTGTTTAGCT CATCCAAAGG
2451 TGTTACCCTG GTTTCATTGA ACCTAACCCC ATTCTTTGTA TCTTCAGTGT
2501 TGGTTTGTTT TAGCTGATCC ATCTGTAACA CAGGAGGGAT CCTTGGCTGA
2551 GGATTGTATT TCAGAACCAC TGACTGCTCT TGACAGTTGT TAACCCACTA
2601 GGCTCCTTTG AGTAGAGAAG CCATAGTCCT TCAGCCTCCA ATTGATATCA
2651 ATACTTAGGA AGACCACAGC TAGACGGACA AACAGCATTG GGAGGCCTTA
```

```
2701 GTCCTGCTCC TTTCAATTCC ATCCTGTAAA GAACAGGAGT CAGGAGCCGC
2751 TGGCAAGAGA CAGCATGTCA CCTGGGACTC TGCCAGTGCA GAATATGAAC
2801 AATGCCATGT TCTTGCAGAA AATGCTTAGC CTGAGTTTCA TAGGAGGTAA
2851 TCACCAGACA ACTGCAGAAT GTAGAACACT GAGCAGGACA ACTGACCTGT
2901 CTCCTTCACA CAGTCCACGT CACCACGAAT CACACAACAA AAAGGAGGAG
2951 AGATATTTTG GGTTCAGAAG AAGTAAATGA TAATGTAGCT ACATTTCTTT
3001 AGTTATTTTG AACCCCAAAT ATTTCCTCAT CTTTTTGTTG TTGTCATTGA
3051 TTTTGGTGAC ATGGACTTGT TTGTAGAGGA CAGGTCAGCT GTCTGGCTCA
3101 ATGGTCTACA TTCTGAAGTT GTCTGAAAAT GTCTTCATGA TTAAATTCAG
3151 CCTAAACGTT TCATCAAGAA CACTACAGAG TCGATACTGT GAGTTTCCAA
3201 CCTCAGCCCA TCTGTGGGCA GAGAAGGTCT AGTTTGTCCA TCAGCATTAT
3251 CATGATATCA GGACTGGTTA CTTGGTTAAG GAGGGGTCTA GGAGATCTGT
3301 CCCTTTTAGA GACACCTTAC TTATGATGAA GTATTTGGGA GAGTGGTTTT
3351 TCAAAGTAGA AATGTCCTGT ATTCCAGTGA TCATCCTCTA AACGTTTTAT
3401 CATTTATTAA TCATCCCTGC CTGTGTCTAT TATTATATTC ATATCTCTAC
3451 GCTGGAAATT TGCTGCCTCA ATGTTTACTG TGCCTTTGTT TTTGCTAGTG
3501 TGTGTTGTTG AAAAAAAAAC ATTCTCTGCC TGAGTTTTAA TTTTTGTCCA
3551 AAGTTATTTT AATCTATACA ATTAAAAACT TTTGCCTATC AAAAAAAAAA
3601 AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 176 bp to 2074 bp; peptide length: 633
Category: similarity to known protein

```
  1 MPLTPTVQGF QWTLRGPDVE TSPFGAPRAA SHGVGRHQEL RDPTVPGPTS
 51 SATNVSMVVS AGPWSGEKAE MNILEINKKS RPQLAENKQQ FRNLKQKCLV
101 TQVAYFLANR QNNYDYEDCK DLIKSMLRDE RLLTEEKLAE ELGQAEELRQ
151 YKVLVHSQER ELTQLREKLQ EGRDASRSLN QHLQALLTPD EPDNSQGRDL
201 REQLAEGCRL AQHLVQKLSP ENDDDEDEDV KVEEAEKVQE LYAPREVQKA
251 EEKEVPEDSL EECAITCSNS HHPCESNQPY GNTRITFEED QVDSTLIDSS
301 SHDEWLDAVC IIPENESDHE QEEEKGPVSP RNLQESEEEE APQESWDEGD
351 WTLSIPPDMS ASYQSDRSTF HSVEEQQVGL ALDIGRHWCD QVKKEDQEAT
401 SPRLSRELLD EKEPEVLQDS LDRFYSTPFE YLELPDLCQP YRSDFYSLQE
451 QHLGLALDLD RMKKDQEEEE DQGPPCPRLS RELPEVVEPE DLQDSLDRWY
501 STPFSYPELP DSCQPYGSCF YSLEEEHVGF SLDVDEIEKY QEGEEDQKPP
551 CPRLNEVLME AEEPEVLQDS LDRCYSTTST YFQLHASFQQ YRSAFYSFEE
601 QDVSLALDVD NRFFTLTVIR HHLAFQMGVI FPH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7d17, frame 2

PIR:T00069 hypothetical protein KIAA0454 - human (fragment), N = 1,
Score = 199, P = 1e-11

PIR:A45592 liver stage antigen LSA-1 - Plasmodium falciparum, N = 1,
Score = 158, P = 2.7e-07

>PIR:T00069 hypothetical protein KIAA0454 - human (fragment)
Length = 1,882

HSPs:

Score = 199 (29.9 bits), Expect = 1.0e-11, P = 1.0e-11
Identities = 74/261 (28%), Positives = 122/261 (46%)

```
Query:   117 EDCKDLIKSMLRDERLLT----EEKLAEELGQAEELRQYKVLVHSQERELTQLREKLQEG 172
             +D + LI+ + + E  L     EEKLAEEL  A     +Y  L+  Q REL+ LR+K++EG
Sbjct:   964 KDLESLIQRVSQLEAQLPKNGLEEKLAEELRSASWPGKYDSLIQDQARELSYLRQKIREG 1023
```

```
Query:   173 RDASRSLNQH-------LQALLTPDEPDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDD 225
             R     + +H        + LL  ++ D   G+  REQLA+G +L + L  KLS ++
Sbjct:  1024 RGICYLITRHAKDTVKSFEDLLRSNDIDYYLGQSFREQLAQGSQLTERLTSKLSTKDHKS 1083

Query:   226 EDEDVKVEEAEKVQELYAPREVQKAEEK-EVPEDSLEECAITCSNSHHPCESNQPYGNTR 284
             E +    +E         L   RE+Q+ E+  EV  +  L+   ++T S+SH   +S++   +T
Sbjct:  1084 EKDQAGLEPLA----LRLSRELQEKEKVIEVLQAKLDARSLTPSSSHALSDSHRSPSSTS 1139

Query:   285 ITFEEDQV--DSTLIDSSSHDEWLDAVCIIPENESDHEQEEEKGPVSPRNLQESEEEEAP 342
                +E +    D  ++    +H E    A     P +                +S +     S  + A
Sbjct:  1140 FLSDELEACSDMDIVSEYTHYEEKKAS---PSHSDSIHHSSHSAVLSSKPSSTSASQGAK 1196

Query:   343 QESWDEGDWTLSIPPDMSASYQSDRSTFH 371
              ES +     +L  P +           S FH
Sbjct:  1197 AES-NSNPISLPTPQNTPKEANQAHSGFH 1224
```

Score = 89 (13.4 bits), Expect = 1.1e-01, P = 1.0e-01
Identities = 35/89 (39%), Positives = 44/89 (49%)

```
Query:   464 KDQEEEEDQG---PPCPRLSRELPEVVEP-EDLQDSLDRWYSTPFSYPELPDSCQ-PYGS 518
             KD + E+DQ     P   RLSREL E  +  E LQ  LD      TP S    L DS + P  +
Sbjct:  1079 KDHKSEKDQAGLEPLALRLSRELQEKEKVIEVLQAKLDARSLTPSSSHALSDSHRSPSST 1138

Query:   519 CFYSLEEEHVGFSLDVDEIEKYQEGEEDQKPP 550
              F S E E        D+D + +Y   EE +  P
Sbjct:  1139 SFLSDELEACS---DMDIVSEYTHYEEKKASP 1167
```

Score = 73 (11.0 bits), Expect = 4.8e+00, P = 9.9e-01
Identities = 31/88 (35%), Positives = 40/88 (45%)

```
Query:   390 DQVKKEDQEATSP---RLSRELLD-EKEPEVLQDSLDRFYSTPFEYLELPDLCQ-PYRSD 444
             D    ++DQ     P   RLSREL + EK  EVLQ  LD     TP      L D  + P  +
Sbjct:  1080 DHKSEKDQAGLEPLALRLSRELQEKEKVIEVLQAKLDARSLTPSSSHALSDSHRSPSSTS 1139

Query:   445 FYSLQEQHLGLALDLDRMKKDQEEEEDQGPP 475
             F S      L     D+D + +      EE +  P
Sbjct:  1140 FLS---DELEACSDMDIVSEYTHYEEKKASP 1167
```

Score = 68 (10.2 bits), Expect = 1.1e-01, P = 1.0e-01
Identities = 36/156 (23%), Positives = 68/156 (43%)

```
Query:    31 SHGVGRHQELRDPTV---PGPTSSATNVSMVVSAGPWS--------GEKAEMNILEINKK 79
             S G  +HQE  + TV   P P S       +  V A                G ++  ++      +
Sbjct:   684 SPGKHQHQEEGNVTVRPFPRPQSLDLGATFTVDAHQLDNQSQPRDPGPQSAFSLPGSTQH 743

Query:    80 SRPQLAENKQQFRNLKQKCLVTQVAYFL-ANRQNNYDYE-DCKDLIKSMLRDERLLTEEK 137
              R QL++ KQ++++L++K L+++    F  AN      Y       + L+K   +   ++  ++
Sbjct:   744 LRSQLSQCKQRYQDLQEKLLLSEATVFAQANELEKYRVMLTGESLVKQDSKQIQVDLQDL 803

Query:   138 LAEELGQAEELRQYKVLVHSQERELTQLREK-LQEG 172
               E   G++E    + +      + E     L+E  L EG
Sbjct:   804 GYETCGRSENEAEREETTSPECEEHNSLKEMVLMEG 839
```

Score = 65 (9.8 bits), Expect = 2.2e-01, P = 2.0e-01
Identities = 23/96 (23%), Positives = 52/96 (54%)

```
Query:   123 IKSMLRDERLLTEEKLAEELGQAEE----LRQYKVLVHSQERELTQLREKLQEGRDASRS 178
             ++   + D+ +   E   + E+     EE      LRQ ++ V   ++ +L +LR+ L        ++ +
Sbjct:     5 LRQRIHDKAVALERAIDEKFSALEEKEKELRQLRLAVRERDHDLERLRDVLS----SNEA 60

Query:   179 LNQHLQALLTPDEPDNSQGRDLREQLAEGCRLAQHLVQKL 218
               Q +++LL          ++G ++ EQL+   C+   Q L +++
Sbjct:    61 TMQSMESLL------RAKGLEV-EQLSTTCQNLQWLKEEM 93
```

Score = 61 (9.2 bits), Expect = 5.5e-01, P = 4.2e-01
Identities = 27/95 (28%), Positives = 47/95 (49%)

```
Query:   134 TEEK-LAEELGQAEELRQY----KVLVHSQERELTQLREKLQEGRDASRSLNQHLQALLT 188
              +E K L   +LG+ EE R Y        +LV       +++    L+  +LQ            ++L      +++L
Sbjct:   855 SERKPLENQLGKQEEFRVYGKSENILV--LRKDIKDLKAQLQNANKVIQNLKSRVRSLSV 912

Query:   189 PDEPDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDDEDE 228
               +   +S   R  R+   A G               ++  SP +   DEDE
Sbjct:   913 TSDYSSSLERP-RKLRAVGT------LEGSSPHSVPDEDE 945
```

Score = 57 (8.6 bits), Expect = 1.4e+00, P = 7.5e-01
Identities = 26/92 (28%), Positives = 47/92 (51%)

```
Query:   127 LRDERLLTEEKLAEELGQAEEL---RQYKVLVHSQERELTQLREKLQEGRDASRSLNQHL 183
             L   E LL   EK+A       Q +E+     R+ ++L+  +     L      R +L E    A R L      L
Sbjct:   358 LTQEVLLLREKVASVESQGQEISGNRRQQLLLMLEG--LVDERSRLNEALQAERQLYSSL 415
```

```
Query:  184 QALLTPDEPDNSQ-GRDLREQLAEGCRLAQHLVQKL 218
                         P++S+  R L+ +L EG ++ +  ++++
Sbjct:  416 VKFHA--HPESSERDRTLQVEL-EGAQVLRSRLEEV 448

 Score = 54 (8.1 bits), Expect = 2.7e+00, P = 9.3e-01
 Identities = 61/264 (23%), Positives = 121/264 (45%)

Query:    3 LTPTVQGFQWTLRGPDVETSPFGAPRAASHGVGRHQE--LRDPTVPGPTSSATNVSMVVS 60
            L+ T Q  QW L+  ++ET  F   +       + +  + L D         SAT   ++
Sbjct:   79 LSTTCQNLQW-LK-EEMETK-FSRWQKEQESIIQQLQTSLHDRNKEVEDLSAT---LLCK 132

Query:   61 AGPWSGEKAEMNILEINKKSR---PQLAENKQQFRNLKQKCLVTQVAYFLANRQNNYDYE 117
             GP   E AE     + +K R     L++  +Q    L+ +  +  +    ++ R+
Sbjct:  133 LGPGQSEIAEELCQRLQRKERMLQDLLSDRNKQV--LEHEMEIQGLLQSVSTREQE-SQA 189

Query:  118 DCKDLIKSMLRDERLLTEEKLAEELGQAEELRQYKVLVHSQERELT---QLREKLQEG-- 172
              + L+++++  ER     + L + LG  + L    +  + +Q+ E+T    +L ++  +G
Sbjct:  190 AAEKLVQALM--ERNSELQALRQYLGGRDSLMS-QAPISNQQAEVTPTGRLGKQTDQGSM 246

Query:  173 RDASRSLNQHLQALLTPDEPDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDDEDEDVKV 232
            +   SR  +  L A       P ++ G DL + +A G         L ++LS  N  +E E +
Sbjct:  247 QIPSRDDSTSLTAKEDVSIPRSTLG-DL-DTVA-G------LEKELS--NAKEELELMAK 295

Query:  233 EEAEKVQELYAPREVQKAEEKEVPEDSLEECAIT 266
            +E E    EL A + +    +E+E+   + +   ++T
Sbjct:  296 KERESQMELSALQSMMAVQEEELQVQAADMESLT 329

 Score = 49 (7.4 bits), Expect = 6.3e+00, P = 1.0e+00
 Identities = 21/87 (24%), Positives = 39/87 (44%)

Query:  192 PDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDDEDEDVKVEEAEKVQELYAPREVQKAE 251
            P ++Q    LR QL++  +  Q L +KL      +        +  E EK + +      + K +
Sbjct:  738 PGSTQ--HLRSQLSQCKQRYQDLQEKLLLS---EATVFAQANELEKYRVMLTGESLVKQD 792

Query:  252 EKEVPEDSLEECAI-TCSNSHHPCESNQ 278
             K++  D L++      TC  S +  E  +
Sbjct:  793 SKQIQVD-LQDLGYETCGRSENEAEREE 819

 Score = 46 (6.9 bits), Expect = 6.3e+00, P = 1.0e+00
 Identities = 19/77 (24%), Positives = 39/77 (50%)

Query:  112 NNYDYEDCKDLIKSMLRDERLLTEEKLAEELGQAEELRQYKVLVHSQERELTQLREKLQ- 170
            + ++ E+ K+  K  +  E ++T+E L+E   QAE  R+ + + +   +  L+E+L
Sbjct:  597 DGWEIEEDKE--KGEVMVETVVTKEGLSESSLQAE-FRKLQGKLKNAHNIINLLKEQLVL 653

Query:  171 EGRDASRSLNQHLQALLT 188
             ++ +  L   L  LT
Sbjct:  654 SSKEGNSKLTPELLVHLT 671
```

Pedant information for DKFZphtes3_7d17, frame 2

Report for DKFZphtes3_7d17.2

[LENGTH]	633
[MW]	72951.15
[pI]	4.40
[HOMOL]	PIR:T00069 hypothetical protein KIAA0454 - human (fragment) 2e-11
[BLOCKS]	BL00201E
[PROSITE]	MYRISTYL 2
[PROSITE]	CK2_PHOSPHO_SITE 14
[PROSITE]	PKC_PHOSPHO_SITE 4
[PROSITE]	ASN_GLYCOSYLATION 2
[PFAM]	TNFR/NGFR cysteine-rich region
[KW]	All_Alpha
[KW]	LOW_COMPLEXITY 4.90 %
[KW]	COILED_COIL 6.95 %

```
SEQ    MPLTPTVQGFQWTLRGPDVETSPFGAPRAASHGVGRHQELRDPTVPGPTSSATNVSMVVS
SEG    ............................................................
PRD    cccccceeeeeeeccccccccccccccccccccccccccccccccccccccceeeeeeeee
COILS  ............................................................

SEQ    AGPWSGEKAEMNILEINKKSRPQLAENKQQFRNLKQKCLVTQVAYFLANRQNNYDYEDCK
SEG    ............................................................
PRD    cccccchhhhhhhheeeccccchhhhhhhhhhcccccchhhhhhhhhccccccccccch
COILS  ............................................................
```

```
SEQ    DLIKSMLRDERLLTEEKLAEELGQAEELRQYKVLVHSQERELTQLREKLQEGRDASRSLN
SEG    ............................................................
PRD    hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhh
COILS  ......................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ    QHLQALLTPDEPDNSQGRDLREQLAEGCRLAQHLVQKLSPENDDDEDEDVKVEEAEKVQE
SEG    ..........................................xxxxxxxxxxxxxxxx..
PRD    hhhhhhhcccccccccchhhhhhhhhhhhhhhhhhhcccccccccchhhhhhhhhhhhhh
COILS  CCCCCCC.....................................................

SEQ    LYAPREVQKAEEKEVPEDSLEECAITCSNSHHPCESNQPYGNTRITFEEDQVDSTLIDSS
SEG    ............................................................
PRD    hhhcchhhhhhhhhhcchhhhhhhccccccccccccccccccceeeeeccccccccccccc
COILS  ............................................................

SEQ    SHDEWLDAVCIIPENESDHEQEEEKGPVSPRNLQESEEEEAPQESWDEGDWTLSIPPDMS
SEG    .................................xxxxxxxxxxxxxxx............
PRD    ccchhhheeeccccccchhhhhccccccccccchhhhhhccccccccccccccccccccc
COILS  ............................................................

SEQ    ASYQSDRSTFHSVEEQQVGLALDIGRHWCDQVKKEDQEATSPRLSRELLDEKEPEVLQDS
SEG    ............................................................
PRD    ccccccccchhhhhhhhhhhhhhhcccccchhhhhcccccchhhhhhhhhhhhheeeecc
COILS  ............................................................

SEQ    LDRFYSTPFEYLELPDLCQPYRSDFYSLQEQHLGLALDLDRMKKDQEEEEDQGPPCPRLS
SEG    ............................................................
PRD    hhhhccceeeeeecccccccccccchhhhhhhhhhhhcchhhhhhhhhhhcccccccccc
COILS  ............................................................

SEQ    RELPEVVEPEDLQDSLDRWYSTPFSYPELPDSCQPYGSCFYSLEEEHVGFSLDVDEIEKY
SEG    ............................................................
PRD    ccceeeeeccchhhhhhhhhcccccccccccccccccccceeeeccceeeccccchhhhhh
COILS  ............................................................

SEQ    QEGEEDQKPPCPRLNEVLMEAEEPEVLQDSLDRCYSTTSTYFQLHASFQQYRSAFYSFEE
SEG    ............................................................
PRD    hcccccccccccchhhhhhhhchhhhhccccceeecceeeehhhhhhhhhhhhhhhhhhc
COILS  ............................................................

SEQ    QDVSLALDVDNRFFTLTVIRHHLAFQMGVIFPH
SEG    .................................
PRD    cchhhhhhcccchhhhhhhhhhhhhhhhhcccc
COILS  .................................
```

Prosite for DKFZphtes3_7d17.2

PS00001	54->58	ASN_GLYCOSYLATION	PDOC00001
PS00001	315->319	ASN_GLYCOSYLATION	PDOC00001
PS00005	13->16	PKC_PHOSPHO_SITE	PDOC00005
PS00005	329->332	PKC_PHOSPHO_SITE	PDOC00005
PS00005	365->368	PKC_PHOSPHO_SITE	PDOC00005
PS00005	401->404	PKC_PHOSPHO_SITE	PDOC00005
PS00006	188->192	CK2_PHOSPHO_SITE	PDOC00006
PS00006	259->263	CK2_PHOSPHO_SITE	PDOC00006
PS00006	286->290	CK2_PHOSPHO_SITE	PDOC00006
PS00006	295->299	CK2_PHOSPHO_SITE	PDOC00006
PS00006	300->304	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006
PS00006	336->340	CK2_PHOSPHO_SITE	PDOC00006
PS00006	345->349	CK2_PHOSPHO_SITE	PDOC00006
PS00006	372->376	CK2_PHOSPHO_SITE	PDOC00006
PS00006	427->431	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	505->509	CK2_PHOSPHO_SITE	PDOC00006
PS00006	522->526	CK2_PHOSPHO_SITE	PDOC00006
PS00006	597->601	CK2_PHOSPHO_SITE	PDOC00006
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	207->213	MYRISTYL	PDOC00008

Pfam for DKFZphtes3_7d17.2

```
HMM_NAME    TNFR/NGFR cysteine-rich region

HMM             *CpeGtYtDWNHvpqClpCtrCePEMGQYMvqPCTwTQNTVC*
                 C+ ++ +  N+ ++      + ++   +  +++ +++ ++VC
Query       274  CESNQPYG-NT-RITFEEDQVDS--TLIDSSSHDEWLDAVC    310
```

DKFZphtes3_7j3
----

group: cell cycle

DKFZphtes3_7j3.2 encodes a novel 628 amino acid putative protein kinase, which is related to the C-TAK1 Cdc25C associated protein kinase.

Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five C associated protein kinase) phosphorylates Cdc25C on serine 216 in vitro. The new protein is closely related to C-Tak1 and therefore should be involved in cell-cycle regulation, too.

The new protein can find application in modulating/blocking the cell cycle.

strong similarity to serine/threonine-specific protein kinases

complete cDNA, complete cds, potential start at Bp 128, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3443 bp
Poly A stretch at pos. 3399, polyadenylation signal at pos. 3376

```
   1 GTGCTTTACT GCGCGCTCTG GTACTGCTGT GGCTCCCCGT CCTGGTGCGG
  51 GACCTGTGCC CCGCGCTTCA GCCCTCCCCG CACAGCCTAC TGATTCCCCT
 101 GCCGCCCTTG CTCACCTCCT GCTCGCCATG GAGTCGCTGG TTTTCGCGCG
 151 GCGCTCCGGC CCCACTCCCT CGGCCGCAGA GCTAGCCCGG CCGCTGGCGG
 201 AAGGGCTGAT CAAGTCGCCC AAGCCCCTAA TGAAGAAGCA GGCGGTGAAG
 251 CGGCACCACC ACAAGCACAA CCTGCGGCAC CGCTACGAGT TCCTGGAGAC
 301 CCTGGGCAAA GGCACCTACG GGAAGGTGAA GAAGGCGCGG GAGAGCTCGG
 351 GGCGCCTGGT GGCCATCAAG TCAATCCGGA AGGACAAAAT CAAAGATGAG
 401 CAAGATCTGA TGCACATACG GAGGGAGATT GAGATCATGT CATCACTCAA
 451 CCACCCTCAC ATCATTGCCA TCCATGAAGT GTTTGAGAAC AGCAGCAAGA
 501 TCGTGATCGT CATGGAGTAT GCCAGCCGGG GCGACCTTTA TGACTACATC
 551 AGCGAGCGGC AGCAGCTCAG TGAGCGCGAA GCTAGGCATT TCTTCCGGCA
 601 GATCGTCTCT GCCGTGCACT ATTGCCATCA GAACAGAGTT GTCCACCGAG
 651 ATCTCAAGCT GGAGAACATC CTCTTGGATG CCAATGGGAA TATCAAGATT
 701 GCTGACTTCG GCCTCTCCAA CCTCTACCAT CAAGGCAAGT TCCTGCAGAC
 751 ATTCTGTGGG AGCCCCCTCT ATGCCTCGCC AGAGATTGTC AATGGGAAGC
 801 CCTACACAGG CCCAGAGGTG GACAGCTGGT CCCTGGGTGT TCTCCTCTAC
 851 ATCCTGGTGC ATGGCACCAT GCCCTTTGAT GGGCATGACC ATAAGATCCT
 901 AGTGAAACAG ATCAGCAACG GGGCCTACCG GGAGCCACCT AAACCCTCTG
 951 ATGCCTGTGG CCTGATCCGG TGGCTGTTGA TGGTGAACCC CACCCGCCGG
1001 GCCACCCTGG AGGATGTGGC CAGTCACTGG TGGGTCAACT GGGGCTACGC
1051 CACCCGAGTG GGAGAGCAGG AGGCTCCGCA TGAGGGTGGG CACCCTGGCA
1101 GTGACTCTGC CCGCGCCTCC ATGGCTGACT GGCTCCGGCG TTCCTCCCGC
1151 CCCCTCCTGG AGAATGGGGC CAAGGTGTGC AGCTTCTTCA AGCAGCATGC
1201 ACCTGGTGGG GGAAGCACCA CCCCTGGCCT GGAGCGCCAG CATTCGCTCA
1251 AGAAGTCCCG CAAGGAGAAT GACATGGCCC AGTCTCTCCA CAGTGACACG
1301 GCTGATGACA CTGCCCATCG CCCTGGCAAG AGCAACCTCA AGCTGCCAAA
1351 GGGCATTCTC AAGAAGAAGG TGTCAGCCTC TGCAGAAGGG GTACAGGAGG
1401 ACCCTCCGGA GCTCAGCCCA ATCCCTGCGA GCCCAGGGCA GGCTGCCCCG
1451 CTGCTCCCCA AGAAGGGCAT TCTCAAGAAG CCCCGACAGC GCGAGTCTGG
1501 CTACTACTCC TCTCCCGAGC CCAGTGAATC TGGGGAGCTC TTGGACGCAG
1551 GCGACGTGTT TGTGAGTGGG GATCCCAAGG AGCAGAAGCC TCCGCAAGCT
1601 TCAGGGCTGC TCCTCCATCG CAAAGGCATC CTCAAACTCA ATGGCAAGTT
1651 CTCCCAGACA GCCTTGGAGC TCGCGGCCCC CACCACCTTC GGCTCCCTGG
1701 ATGAACTCGC CCCACCTCGC CCCCTGGCCC GGGCCAGCCG ACCCTCAGGG
1751 GCTGTGAGCG AGGACAGCAT CCTGTCCTCT GAGTCCTTTG ACCAGCTGGA
1801 CTTGCCTGAA CGGCTCCCAG AGCCCCCACT GCGGGGCTGT GTGTCTGTGG
1851 ACAACCTCAC GGGGCTTGAG GAGCCCCCCT CAGAGGGCCC TGGAAGCTGC
1901 CTGAGGCGCT GGCGGCAGGA TCCTTTGGGG GACAGCTGCT TTTCCCTGAC
1951 AGACTGCCAG GAGGTGACAG CGACCTACCG ACAGGCACTG AGGGTCTGCT
2001 CAAAGCTCAC CTGAGTGGAG TAGGCATTGC CCCAGCCCGG TCAGGCTCTC
2051 AGATGCAGCT GGTTGCACCC CGAGGGGAGA TGCCTTCTCC CCCACCTCCC
2101 AGGACCTGCA TCCCAGCTCA GAAGGCTGAG AGGGTTTGCA GTGGAGCCCT
2151 GAGCAGGGCT GGATATGGGA AGTAGGCAAA TGAAATGCGC CAAGGGTTCA
2201 GTGTCTGTCT TCAGCCCTGC TGAACGAAGA GGATACTAAA GAGAGGGGAA
2251 CGGGAATGCC CGCGACAGAG TCCACATTGC CTGTTTCTTG TGTACATGGG
2301 GGGGCCACAG AGACCTGGAA AGAGAACTCT CCCAGGGCCC ATCTCCTGCA
2351 TCCCATGAAT ACTCTGTACA CATGGTGCCT TCTAAGGACA GCTCCTTCCC
2401 TACTCATTCC CTGCCCAAGT GGGGCCAGAC CTCTTTACAC ACACATTCCC
2451 GTTCCTACCA ACCACCAGAA CTGGATGGTG GCACCCCTAA TGTGCATGAG
2501 GCATCCTGGG AATGGTCTGG AGTAACGCTT CGTTATTTTT ATTTTTATTT
```

```
2551 TTATTTATTT ATTTATTTTT TTGAGACGGA GTTTCGCTCT TGGTGCCCAG
2601 GCTAGAGTGC AATGGCGCGA TCTCAGCTCA CCTCAACCTC CGCCTCCCGG
2651 GTTCAAGCGA TTCTCCTGCC TCAGCCTCCC TAGTAGCTGG GATTACAGGC
2701 GCCCGCCACC ATGCCCGGCT AATTTTGTAT TTTTAGTAGA GACAGGGTTT
2751 CTCCATGTTG GTCAGGCTGG TCTCAAACTC CCGACCTCAG GTGATCCACC
2801 CACCTCGGCC TCCCAAAGTG CTGGGATTAC AGGCGTGAGC CACCGCGCCC
2851 CACCTAACCC TTCCTTATTT AGCCTAGGAG TAAGAGAACA CAATCTCTGT
2901 TTCTTCAATG GTTCTCTTCC CTTTTCCATC CTCCAAACCT GGCCTGAGCC
2951 TCCTGAAGTT GCTGCTGTGA ATCTGAAAGA CTTGAAAAGC CTCCGCCTGC
3001 TGTGTGGACT TCATCTCAAG GGGCCCAGCC TCCTCTGGAC TCCACCTTGG
3051 ACCTCAGTGA CTCAGAACTT CTGCCTCTAA GCTGCTCTAA AGTCCAGACT
3101 ATGGATGTGT TCTCTAGGCC TTCAGGACTC TAGAATGTCC ATATTTATTT
3151 TTATGTTCTT GGCTTTGTGT TTTAGGAAAA GTGAATCTTG CTGTTTTCAA
3201 TAATGTGAAT GCTATGTTCT GGGAAAATCC ACTATGACAT CTAAGTTTTG
3251 TGTACAGAGA GATATTTTTG CAACTATTTC CACCTCCTCC CACAACCCCC
3301 CACACTCCAC TCCACACTCT TGAGTCTCTT TACCTAATGG TCTCTACCTA
3351 ATGGACCTCC GTGGCCAAAA AGTACCATTA AAACCAGAAA GGTGATTGGA
3401 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

98202387:
C-TAK1 protein kinase phosphorylates human Cdc25C on serine 216 and promotes 14-3-3
protein binding.

Peptide information for frame 2

ORF from 128 bp to 2011 bp; peptide length: 628
Category: strong similarity to known protein

```
  1 MESLVFARRS GPTPSAAELA RPLAEGLIKS PKPLMKKQAV KRHHHKHNLR
 51 HRYEFLETLG KGTYGKVKKA RESSGRLVAI KSIRKDKIKD EQDLMHIRRE
101 IEIMSSLNHP HIIAIHEVFE NSSKIVIVME YASRGDLYDY ISERQQLSER
151 EARHFFRQIV SAVHYCHQNR VVHRDLKLEN ILLDANGNIK IADFGLSNLY
201 HQGKFLQTFC GSPLYASPEI VNGKPYTGPE VDSWSLGVLL YILVHGTMPF
251 DGHDHKILVK QISNGAYREP PKPSDACGLI RWLLMVNPTR RATLEDVASH
301 WWVNWGYATR VGEQEAPHEG GHPGSDSARA SMADWLRRSS RPLLENGAKV
351 CSFFKQHAPG GGSTTPGLER QHSLKKSRKE NDMAQSLHSD TADDTAHRPG
401 KSNLKLPKGI LKKKVSASAE GVQEDPPELS PIPASPGQAA PLLPKKGILK
451 KPRQRESGYY SSPEPSESGE LLDAGDVFVS GDPKEQKPPQ ASGLLLHRKG
501 ILKLNGKFSQ TALELAAPTT FGSLDELAPP RPLARASRPS GAVSEDSILS
551 SESFDQLDLP ERLPEPPLRG CVSVDNLTGL EEPPSEGPGS CLRRWRQDPL
601 GDSCFSLTDC QEVTATYRQA LRVCSKLT
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7j3, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_7j3, frame 2

Report for DKFZphtes3_7j3.2

[LENGTH] 628
[MW] 69612.39
[pI] 9.01
[HOMOL] TREMBL:AB011109_1 gene: "KIAA0537"; product: "KIAA0537 protein"; Homo sapiens mRNA for KIAA0537 protein, complete cds. 1e-152
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w] 5e-66
[FUNCAT] 11.01 stress response [S. cerevisiae, YDR477w] 5e-66

[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 5e-66
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YLR096w] 6e-54
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YLR096w] 6e-54
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c] 8e-52
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 8e-52
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 9e-51
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL101w] 9e-51
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 1e-45
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 6e-44
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 2e-42
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-34
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL139w CTK1 - carboxy-terminal domain] 2e-28
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 4e-28
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YGL180w] 2e-26
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 2e-26
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YGL180w] 2e-26
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YER129w] 4e-26
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae, YPL031c] 5e-24
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c] 5e-24
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YHL007c] 6e-24
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YHL007c] 6e-24
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNR031c] 1e-22
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YNR031c] 1e-22
[FUNCAT] 03.13 meiosis [S. cerevisiae, YDR523c] 8e-22
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w] 6e-21
[FUNCAT] 06.07 protein modification (glycolsylation, acylation, myristylation, palmitylation, farnesylation and processing) [S. cerevisiae, YFL033c] 6e-21
[FUNCAT] 10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 7e-19
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YDL159w] 3e-18
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae, YNL183c] 1e-17
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c] 1e-17
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 2e-17
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YNL020c] 4e-16
[FUNCAT] 04.03.99 other trna-transcription activities [S. cerevisiae, YOR061w] 1e-15
[FUNCAT] 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w] 5e-15
[FUNCAT] c energy conversion [M. genitalium, MG109] 3e-12
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YBR097w] 2e-08
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c] 8e-05
[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YHR079c] 8e-05
[BLOCKS] BL00479C Phorbol esters / diacylglycerol binding domain proteins
[BLOCKS] BL00239B Receptor tyrosine kinase class II proteins
[BLOCKS] BL00107A Protein kinases ATP-binding region proteins
[SCOP] d1gol__ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus 1e-77
[SCOP] d1wfc__ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens) 4e-68
[SCOP] d1koa_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditi 2e-85
[SCOP] d1koba_ 5.1.1.1.6 Twitchin, kinase domain [california sea har 1e-80
[SCOP] d1phk__ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinas 2e-76
[SCOP] d1irk__ 5.1.1.2.4 insulin receptor [Human (Homo sapiens) 1e-69
[SCOP] d1apme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu 1e-84
[SCOP] d1fgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Hom 1e-68
[SCOP] d1ydre_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo 9e-85
[SCOP] d1fmk_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Hom 1e-69
[SCOP] d1cdka_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su 1e-85
[SCOP] d2hcka3 5.1.1.2.1 (167-437) Haemopoetic cell kinase Hck [huma 5e-66
[SCOP] d1csn__ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe 9e-47
[SCOP] d1jsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens) 1e-75
[SCOP] d1ckja_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus) 5e-54
[EC] 2.7.1.38 Phosphorylase kinase 1e-36
[EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 4e-40

[EC]	2.7.1.128 [Acetyl-CoA carboxylase] kinase 1e-61
[EC]	2.7.1.117 Myosin-light-chain kinase 2e-40
[EC]	2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 1e-61
[EC]	2.7.1.37 Protein kinase 7e-42
[PIRKW]	phosphotransferase 6e-66
[PIRKW]	nucleus 1e-64
[PIRKW]	calcium 7e-35
[PIRKW]	duplication 1e-38
[PIRKW]	tandem repeat 4e-39
[PIRKW]	phorbol ester binding 1e-38
[PIRKW]	zinc 1e-38
[PIRKW]	cell cycle control 1e-42
[PIRKW]	serine/threonine-specific protein kinase 8e-68
[PIRKW]	oncogene 1e-40
[PIRKW]	phospholipid binding 1e-38
[PIRKW]	autophosphorylation 1e-64
[PIRKW]	brain 1e-40
[PIRKW]	heterotetramer 2e-36
[PIRKW]	mitosis 7e-42
[PIRKW]	polymer 1e-35
[PIRKW]	magnesium 6e-66
[PIRKW]	ATP 8e-68
[PIRKW]	polyprotein 1e-40
[PIRKW]	phosphoprotein 1e-64
[PIRKW]	apoptosis 4e-39
[PIRKW]	glycoprotein 7e-42
[PIRKW]	leucine zipper 3e-35
[PIRKW]	skeletal muscle 7e-35
[PIRKW]	protein kinase 5e-41
[PIRKW]	cAMP binding 3e-38
[PIRKW]	testis 9e-36
[PIRKW]	purine nucleotide binding 2e-49
[PIRKW]	calcium binding 8e-39
[PIRKW]	alternative splicing 3e-37
[PIRKW]	P-loop 2e-49
[PIRKW]	lipoprotein 2e-33
[PIRKW]	segmentation 1e-33
[PIRKW]	core protein 1e-40
[PIRKW]	muscle 7e-35
[PIRKW]	myristylation 2e-33
[PIRKW]	EF hand 8e-39
[PIRKW]	cell division 2e-40
[PIRKW]	calmodulin binding 4e-40
[SUPFAM]	ribosomal protein S6 kinase II 5e-36
[SUPFAM]	fibronectin type III repeat homology 3e-33
[SUPFAM]	immunoglobulin homology 3e-33
[SUPFAM]	calcium-dependent protein kinase 8e-39
[SUPFAM]	AMP-activated protein kinase 6e-66
[SUPFAM]	protein kinase akt 3e-42
[SUPFAM]	protein kinase SPK1 1e-42
[SUPFAM]	unassigned Ser/Thr or Tyr-specific protein kinases 8e-68
[SUPFAM]	Ca2+/calmodulin-dependent protein kinase 3e-37
[SUPFAM]	calmodulin repeat homology 8e-39
[SUPFAM]	cAMP receptor protein cyclic nucleotide-binding domain homology 6e-33
[SUPFAM]	protein kinase C zeta 1e-36
[SUPFAM]	Dictyostelium cAMP-dependent protein kinase catalytic chain 1e-34
[SUPFAM]	death-associated protein kinase 4e-39
[SUPFAM]	pleckstrin repeat homology 3e-42
[SUPFAM]	ankyrin repeat homology 4e-39
[SUPFAM]	protein kinase homology 8e-68
[SUPFAM]	Ca2+/calmodulin-dependent protein kinase II 8e-41
[SUPFAM]	protein kinase C zinc-binding repeat homology 1e-38
[SUPFAM]	twitchin 3e-33
[SUPFAM]	protein kinase C delta 1e-38
[SUPFAM]	cGMP-dependent protein kinase 6e-33
[SUPFAM]	protein kinase cdr1 7e-42
[SUPFAM]	protein kinase C C2 region homology 3e-37
[SUPFAM]	protein kinase C alpha 3e-37
[SUPFAM]	yeast protein kinase C 5e-36
[SUPFAM]	kinase-related transforming protein 1e-41
[SUPFAM]	kinase interaction domain homology 1e-42
[SUPFAM]	gag-akt polyprotein 1e-40
[SUPFAM]	Ca2+/calmodulin-dependent protein kinase I 4e-40
[SUPFAM]	protein kinase C mu 4e-33
[PROSITE]	PROTEIN_KINASE_ATP 2
[PROSITE]	RGD 1
[PROSITE]	MYRISTYL 4
[PROSITE]	CAMP_PHOSPHO_SITE 3
[PROSITE]	CK2_PHOSPHO_SITE 13
[PROSITE]	TYR_PHOSPHO_SITE 2
[PROSITE]	PKC_PHOSPHO_SITE 12

```
[PROSITE]     ASN_GLYCOSYLATION     2
[PROSITE]     PROTEIN_KINASE_ST     1
[PFAM]        Eukaryotic protein kinase domain
[KW]          All_Alpha
[KW]          3D
[KW]          LOW_COMPLEXITY    10.51 %

SEQ    MESLVFARRSGPTPSAAELARPLAEGLIKSPKPLMKKQAVKRHHHKHNLRHRYEFLETLG
SEG    ........................................xxxxxxxxxxxx........
1ctpE  ........................HHHHHHHHHHHHHHCCCCCCCC--GGGEEEEEEEEE

SEQ    KGTYGKVKKARESSGRLVAIKSIRKDKIKDEQDLMHIRREIEIMSSLNHPHIIAIHEVFE
SEG    ............................................................
1ctpE  CTTTEEEEEEEETTTEEEEEEEEEEHHHHHHCCHHHHHHHHHHHHCCCTTTBCCEEEEEE

SEQ    NSSKIVIVMEYASRGDLYDYISERQQLSEREARHFFRQIVSAVHYCHQNRVVHRDLKLEN
SEG    ............................................................
1ctpE  ETTEEEEEEEECTTTTBHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHCCEECCCCCGGG

SEQ    ILLDANGNIKIADFGLSNLYHQGKFLQTFCGSPLYASPEIVNGKPYTGPEVDSWSLGVLL
SEG    ............................................................
1ctpE  EEETTTTCEEECCTTTTEET-TTT-BCCCCCCGGGCCHHHHHCCCBC-HHHHHHHHHHH

SEQ    YILVHGTMPFDGHDHKILVKQISNGAYREPPKPSDACGLIRWLLMVNPTRRATLEDVASH
SEG    ............................................................
1ctpE  HHHHHCCTTTTTTTHHHHHHHHHHCCCCCTTTCHHHHHHHHHTTTTTGGGTTTHHHHHHC

SEQ    WWVNWGYATRVGEQEAPHEGGHPGSDSARASMADWLRRSSRPLLENGAKVCSFFKQHAPG
SEG    ............................................................
1ctpE  GG..........................................................

SEQ    GGSTTPGLERQHSLKKSRKENDMAQSLHSDTADDTAHRPGKSNLKLPKGILKKKVSASAE
SEG    ............................................................
1ctpE  ............................................................

SEQ    GVQEDPPELSPIPASPGQAAPLLPKKGILKKPRQRESGYYSSPEPSESGELLDAGDVFVS
SEG    ....................xxxxxxxxxxxx...xxxxxxxxxxxxxxx..........
1ctpE  ............................................................

SEQ    GDPKEQKPPQASGLLLHRKGILKLNGKFSQTALELAAPTTFGSLDELAPPRPLARASRPS
SEG    ..............................................xxxxxxxxxxxxxx
1ctpE  ............................................................

SEQ    GAVSEDSILSSESFDQLDLPERLPEPPLRGCVSVDNLTGLEEPPSEGPGSCLRRWRQDPL
SEG    ................xxxxxxxxxxxxx...............................
1ctpE  ............................................................

SEQ    GDSCFSLTDCQEVTATYRQALRVCSKLT
SEG    ............................
1ctpE  ............................
```

Prosite for DKFZphtes3_7j3.2

PS00001	121->125	ASN_GLYCOSYLATION	PDOC00001
PS00001	576->580	ASN_GLYCOSYLATION	PDOC00001
PS00004	290->294	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	337->341	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	413->417	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	74->77	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	142->145	PKC_PHOSPHO_SITE	PDOC00005
PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	289->292	PKC_PHOSPHO_SITE	PDOC00005
PS00005	327->330	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	373->376	PKC_PHOSPHO_SITE	PDOC00005
PS00005	377->380	PKC_PHOSPHO_SITE	PDOC00005
PS00005	616->619	PKC_PHOSPHO_SITE	PDOC00005
PS00006	15->19	CK2_PHOSPHO_SITE	PDOC00006
PS00006	133->137	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	227->231	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	331->335	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	391->395	CK2_PHOSPHO_SITE	PDOC00006

PS00006	461->465	CK2_PHOSPHO_SITE	PDOC00006
PS00006	511->515	CK2_PHOSPHO_SITE	PDOC00006
PS00006	523->527	CK2_PHOSPHO_SITE	PDOC00006
PS00006	578->582	CK2_PHOSPHO_SITE	PDOC00006
PS00006	606->610	CK2_PHOSPHO_SITE	PDOC00006
PS00007	453->460	TYR_PHOSPHO_SITE	PDOC00007
PS00007	453->461	TYR_PHOSPHO_SITE	PDOC00007
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	324->330	MYRISTYL	PDOC00008
PS00008	347->353	MYRISTYL	PDOC00008
PS00008	360->366	MYRISTYL	PDOC00008
PS00016	134->137	RGD	PDOC00016
PS00107	59->82	PROTEIN_KINASE_ATP	PDOC00100
PS00107	59->86	PROTEIN_KINASE_ATP	PDOC00100
PS00108	171->184	PROTEIN_KINASE_ST	PDOC00100

Pfam for DKFZphtes3_7j3.2

```
HMM_NAME      Eukaryotic protein kinase domain

HMM                *YeigRiIGeGsFGtVYkCiWrTGeIVAIKIIkkrsms.......FlREI
                    YE+++++G+G++G+V+K+++ +G++VAIK I+K++++       ++REI
Query         53    YEFLETLGKGTYGKVKKARESSGRLVAIKSIRKDKIKDEQDLMHIRREI   101

HMM                qIMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEw
                   +IM +LNHP+II + ++FE  ++ I ++MEY+  GDL+DYI+++ ++SE+
Query        102   EIMSSLNHPHIIAIHEVFE-NSSKIVIVMEYASRGDLYDYISERQQLSER   150

HMM                eIrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeNgqIKIcDFGLARqM
                   E+R++++QI++++ Y+H ++++HRDLK ENIL+D NG+IKI+DFGL+ ++
Query        151   EARHFFRQIVSAVHYCHQNRVVHRDLKLENILLDANGNIKIADFGLSNLY   200

HMM                nnYerMttfCGTPWYMMAPEVIImg.nyYttkVDMWSFGCILWEMMTGep
                   + + ++ TFCG+P Y  +PE+ ++G +Y +++VD WS+G++L++++ G+
Query        201   HQGKFLQTFCGSPLYA-SPEI-VNGKPYTGPEVDSWSLGVLLYILVHGTM   248

HMM                PFyddnMemImrIiqrfrrpfWpnCSeElyDFMrwCWnyDPekRPTFrQI
                   PF+++    ++  I + +++ +P  S+ +  ++RW++ ++P++R T +++
Query        249   PFDGHDHKILVKQISNGAYREPPKPSD-ACGLIRWLLMVNPTRRATLEDV   297

HMM                LnHPWF*
                     H W+
Query        298   ASHWWV    303
```

DKFZphtes3_7j8
----

group: testes derived

DKFZphtes3_7j8 encodes a novel 410 amino acid protein nearly identical to human WUGSC:H_DJ1159O04.1.

The novel protein contains an additional C-terminal domain, which is not present in WUGSC:H_DJ1159O04.1.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

WUGSC:H_DJ1159O04.1 similarity to YBL104p

verifies and extends the genmodel WUGSC:H_DJ1159O04.1
similarity to S.cerevisiae YBL104p

Sequenced by BMFZ

Locus: /map="7p21-p22"

Insert length: 3353 bp
Poly A stretch at pos. 3231, no polyadenylation signal found

```
   1 GCAAAATATG TTGTATTTGT GGCATAGTTC ATATTTACAC TATCATAAAA
  51 TTATGGCCGA GAAGTTAAAT ATTCTAAATG TGTCAACATA GTTCTCTGTA
 101 AAACTGACTT ATTTTCCAAA TATATTTTGA AATAAAACAA TATAAAAATG
 151 TTTTCTGTTT TTAGGAATGG TGGAAAGCAG CAGACATAAT TGGAGTGGGT
 201 TGGATAAGCA AAGTGATATT CAAAATTTAA ATGAAGAGAG AATCTTAGCT
 251 TTACAGCTTT GTGGGTGGAT AAAGAAAGGA ACGGATGTAG ACGTGGGGCC
 301 ATTTTTGAAC TCCCTTGTAC AAGAAGGGGA ATGGGAAAGA GCTGCTGCTG
 351 TGGCATTGTT CAACTTGGAT ATTCGCCGAG CAATCCAAAT CCTGAATGAA
 401 GGGGCATCTT CTGAAAAAGG AGATCTGAAT CTCAATGTGG TAGCAATGGC
 451 TTTATCGGGT TATACGGATG AGAAGAACTC CCTTTGGAGA GAAATGTGTA
 501 GCACACTGCG ATTACAGCTA AATAACCCGT ATTTGTGTGT CATGTTTGCA
 551 TTTCTGACAA GTGAAACAGG ATCTTACGAT GGAGTTTTGT ATGAAAACAA
 601 AGTTGCAGTA CGTGACAGAG TGGCATTTGC TTGTAAATTC CTTAGTGATA
 651 CTCAGTTAAA TAGATACATC GAAAAGTTGA CCAATGAAAT GAAAGAGGCT
 701 GGAAATTTGG AAGGAATTTT GCTTACAGGC CTTACTAAAG ATGGAGTGGA
 751 CTTAATGGAG AGTTATGTTG ATAGAACTGG AGATGTTCAA ACAGCAAGTT
 801 ACTGTATGTT ACAGGGTTCA CCTTTAGATG TTCTTAAAGA TGAAAGGGTT
 851 CAGTACTGGA TTGAGAATTA TAGAAATTTA TTAGATGCCT GGAGGTTTTG
 901 GCATAAACGA GCTGAATTTG ATATTCACAG GAGTAAGTTG GATCCCAGTT
 951 CCAAGCCTTT AGCACAAGTT TTTGTGAGTT GCAATTTCTG TGGCAAGTCA
1001 ATCTCCTACA GCTGTTCAGC TGTGCCTCAT CAGGGCAGAG GTTTTAGTCA
1051 GTATGGTGTG AGTGGCTCAC CAACGAAATC TAAAGTCACA AGTTGTCCTG
1101 GCTGTCGAAA ACCACTTCCT CGATGTGCGC TTTGTCTCAT TAATATGGGA
1151 ACACCAGTTT CTAGCTGTCC TGGAGGAACC AAATCAGATG AAAAAGTGGA
1201 CTTGAGCAAG GACAAAAAAT TAGCCCAATT TAACAACTGG TTTACATGGT
1251 GTCATAATTG CAGGCACGGT GGACATGCTG GACATATGCT TAGTTGGTTC
1301 AGGGACCATG CAGAGTGCCC TGTGTCTGCA TGCACGTGTA AATGTATGCA
1351 GTTGGATACA ACGGGGAATC TGGTACCTGC AGAGACTGTC CAGCCATAAA
1401 ATGTTACCAC CTTAAGAGAA CCCTTCAAGT GTGGAGCTTT CTAGTAGGTG
1451 TCCTTCATAG CTCAGAAACA TACCTCAGAA CAAGCCATTC ATGACTTACC
1501 TGTAATGGGA AAATAAATCA TTCTATCAGA TCAGCAGTTT TGATGTTTGA
1551 GTGATTTTGA TATGCTTCAC AGAGACAAAT GCTGCCAAAA TAAACATCGA
1601 AGTATAGACA TGAGTTCTGT TCAGCAGGTT GAAAAGTCTG ATTTAGAAAA
1651 ACTTTCTAAG TTTTGGTTGA AATTATGAAC ACTCTAGAAG CAGAATTTCT
1701 GGAAGAGCCA AGAACAGACT TTGAGCCTAT ATCTTCAAAG CTGAAACTGG
1751 ATATCTTTCA ATAAAATATG TGCACTTTTA AAATAAAATG ACTAATTCTG
1801 TGATTCAGAC AATAGTTTTA AGTTCAGCTG TGCTTAGATT TCTTTCAGAT
1851 TAATTTAAAA TTATAGATTT TTACTTTTAG AATTGCAGAG CCCCTATCCC
1901 ACACTGGAGA ATATTTTTTA TTACTGTCTG TTATATATGT GTCTATGTGT
1951 GTGTGTATAT TTATGTGTGT ATGTATAAAT ATGTACTTTT TAAAGGAGCC
2001 TTTTCCCTCC TTTGATTTTA AGATAAGCAA TCTTTTGGCA TAACATTATC
2051 GTCTTCCTAG AAAAGCCAAG ATGAAGAATC TATCTTACAA CTTTTTCTCT
2101 TCAGTAGAGA AAAACATGTA CCATTTCAGG TGAACATACA AAATTTTCAC
2151 TTTCTACCTT TTGCCTTCCA ATGTCCTGAT TTGTCTTCAA AGGTTTTTCT
2201 CCATATTAAT TTGTCATCTT ATCCTCATCA CCTGAGAACA TTTTACTGCA
2251 TACAAAGTCT ATGCAAGATT ATATGTAACT AGCCATTTAG TATAATCTAT
2301 GTCAGTGTTT CTGTGCTGTC AAATTCCGTC CTGATTTGGA ATACCATACC
2351 TTGTTCTTTC CAAGGTAGAC TAGGAAGTGT TGGGGAAATA GGGTCACTTC
2401 AGAGACCATT TTAGATGTAA GTTTTTAAAT GTAAGTGTTA CTGGGGCTAA
2451 GTCAGGGACT TTATTTAAAA CATTTTTTTT TTCTCATTTC ATAGCTAGAT
2501 AGTTGTAAGA GAAATACAAA GAATTTACAA GATGCTTCTC TGTCATCTGC
```

```
2551 CGTATGCAGA GGGACTGAAC TAGGAATTTT GTAGTTGAAG CTGTGTTCAT
2601 AAAGAGTAAA TCTTATTTTA TAGATTTTGG AGAAATAAAA CAAGAATTTT
2651 AAGAGCTTTC GTATTAGCAG TTTTGCCTTA TAAAAACTAA GATTTGTCAG
2701 ATTAGTTTGA GGTGTAACCT AAATATTAAA AGTAGATTAA ATTTATTTTT
2751 TACCTTGAGT GTCTGATACA TAAAACCCTT TTCTAGGAAA ACATTGGAAG
2801 TAGTACATAT TTACTCTAAA TGTCTCACCT GCATGACAGT CTTTTCAAAT
2851 GAAAGACATG GTAATTGCAA TTTTTTTTTA AAGATTGCTA TTAAGGGTAC
2901 TTTTTCCAGC CTTCATTTGA GTAAATCTTA ATTGATTTCA TTTTATTAAC
2951 ATATACCCTT TACCTTTAAT ATTTCATTTG AAGTGTTCCT TTCAAACTTA
3001 CTGTCTTAAA TATGAAAGTC AGCTTTAAGT AATGTCAGAC TCATATGCAT
3051 TTTCATTCTC ATTAGCTAAA GTAAAATGTA AAATTATCTC AAATAGTTAC
3101 AAGTTTTGGA AATACAGTAT AAAACATGAA TGTAAAGTCT ATTATGTAAT
3151 ATGCTTATTT GTAATCCTAA TATATGAGGG TGACATTTTT AAGATTGTAT
3201 GTATGTGTCA ACCTCTTAAA TGTTTTCTGT GAAAAAAAAA AAAAAAAAAA
3251 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3301 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 167 bp to 1396 bp; peptide length: 410
Category: known protein
Classification: unclassified

```
  1 MVESSRHNWS GLDKQSDIQN LNEERILALQ LCGWIKKGTD VDVGPFLNSL
 51 VQEGEWERAA AVALFNLDIR RAIQILNEGA SSEKGDLNLN VVAMALSGYT
101 DEKNSLWREM CSTLRLQLNN PYLCVMFAFL TSETGSYDGV LYENKVAVRD
151 RVAFACKFLS DTQLNRYIEK LTNEMKEAGN LEGILLTGLT KDGVDLMESY
201 VDRTGDVQTA SYCMLQGSPL DVLKDERVQY WIENYRNLLD AWRFWHKRAE
251 FDIHRSKLDP SSKPLAQVFV SCNFCGKSIS YSCSAVPHQG RGFSQYGVSG
301 SPTKSKVTSC PGCRKPLPRC ALCLINMGTP VSSCPGGTKS DEKVDLSKDK
351 KLAQFNNWFT WCHNCRHGGH AGHMLSWFRD HAECPVSACT CKCMQLDTTG
401 NLVPAETVQP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7j8, frame 2

PIR:S45391 probable membrane protein YBL104c - yeast (Saccharomyces cerevisiae), N = 2, Score = 446, P = 4.5e-47

TREMBL:AC004982_1 gene: "WUGSC:H_DJ1159O04.1"; Homo sapiens PAC clone DJ1159O04 from 7p21-p22, complete sequence., N = 1, Score = 2038, P = 7.6e-211

>TREMBL:AC004982_1 gene: "WUGSC:H_DJ1159O04.1"; Homo sapiens PAC clone DJ1159O04 from 7p21-p22, complete sequence.
Length = 379

HSPs:

Score = 2038 (305.8 bits), Expect = 7.6e-211, P = 7.6e-211
Identities = 379/379 (100%), Positives = 379/379 (100%)

```
Query:    1 MVESSRHNWSGLDKQSDIQNLNEERILALQLCGWIKKGTDVDVGPFLNSLVQEGEWERAA 60
            MVESSRHNWSGLDKQSDIQNLNEERILALQLCGWIKKGTDVDVGPFLNSLVQEGEWERAA
Sbjct:    1 MVESSRHNWSGLDKQSDIQNLNEERILALQLCGWIKKGTDVDVGPFLNSLVQEGEWERAA 60

Query:   61 AVALFNLDIRRAIQILNEGASSEKGDLNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN 120
            AVALFNLDIRRAIQILNEGASSEKGDLNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN
Sbjct:   61 AVALFNLDIRRAIQILNEGASSEKGDLNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN 120
```

```
Query:  121 PYLCVMFAFLTSETGSYDGVLYENKVAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN 180
            PYLCVMFAFLTSETGSYDGVLYENKVAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN
Sbjct:  121 PYLCVMFAFLTSETGSYDGVLYENKVAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN 180

Query:  181 LEGILLTGLTKDGVDLMESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD 240
            LEGILLTGLTKDGVDLMESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD
Sbjct:  181 LEGILLTGLTKDGVDLMESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD 240

Query:  241 AWRFWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSISYSCSAVPHQGRGFSQYGVSG 300
            AWRFWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSISYSCSAVPHQGRGFSQYGVSG
Sbjct:  241 AWRFWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSISYSCSAVPHQGRGFSQYGVSG 300

Query:  301 SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT 360
            SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT
Sbjct:  301 SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT 360

Query:  361 WCHNCRHGGHAGHMLSWFR 379
            WCHNCRHGGHAGHMLSWFR
Sbjct:  361 WCHNCRHGGHAGHMLSWFR 379
```

Pedant information for DKFZphtes3_7j8, frame 2

Report for DKFZphtes3_7j8.2

```
[LENGTH]        410
[MW]            45862.45
[pI]            6.51
[HOMOL]         TREMBL:AC004982_1 gene: "WUGSC:H_DJ1159O04.1"; Homo sapiens PAC clone DJ1159O04
from 7p21-p22, complete sequence. 0.0
[FUNCAT]        99 unclassified proteins       [S. cerevisiae, YBL104c] 7e-48
[BLOCKS]        BL00028 Zinc finger, C2H2 type, domain proteins
[BLOCKS]        BL00534A Ferrochelatase proteins
[PIRKW]         transmembrane protein 2e-46
[KW]            All_Alpha
```

```
SEQ   MVESSRHNWSGLDKQSDIQNLNEERILALQLCGWIKKGTDVDVGPFLNSLVQEGEWERAA
PRD   cccccccccccccccccchhhhhhhhhhhhhhcccccccccccccccccccccccchhhhh

SEQ   AVALFNLDIRRAIQILNEGASSEKGDLNLNVVAMALSGYTDEKNSLWREMCSTLRLQLNN
PRD   hhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhccc

SEQ   PYLCVMFAFLTSETGSYDGVLYENKVAVRDRVAFACKFLSDTQLNRYIEKLTNEMKEAGN
PRD   cccccceeeccccccccccccceeeccchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhcc

SEQ   LEGILLTGLTKDGVDLMESYVDRTGDVQTASYCMLQGSPLDVLKDERVQYWIENYRNLLD
PRD   cceeeeeecccchhhhhhhhcccccceeeeecccccccccccchhhhhhhhhhhhhhh

SEQ   AWRFWHKRAEFDIHRSKLDPSSKPLAQVFVSCNFCGKSISYSCSAVPHQGRGFSQYGVSG
PRD   hhhhhhhhhhhhhhcccccccccceeeeeeeccccccccccccccccccccccccccccc

SEQ   SPTKSKVTSCPGCRKPLPRCALCLINMGTPVSSCPGGTKSDEKVDLSKDKKLAQFNNWFT
PRD   ccccccccccccccccccceeeeecccccccccccccccccccceeeehhhhhhhhhcceee

SEQ   WCHNCRHGGHAGHMLSWFRDHAECPVSACTCKCMQLDTTGNLVPAETVQP
PRD   eeccccccccccchhhhhhhhhcccccccccccccccccccccccccccc
```

(No Prosite data available for DKFZphtes3_7j8.2)

(No Pfam data available for DKFZphtes3_7j8.2)

DKFZphtes3_7p10
----

group: Cell Cycle

DKFZphtes3_7p10.1 encodes a novel 422 amino acid putative protein, which is closely related to the Xenopus laevis XPMC2 protein.

In fission yeast the kinases Weel and Mikl control that initiation of mitosis starts after completion of DNA synthesis. Yeast in which both Weel and Mikl kinases are defective exhibit a mitotic catastrophe phenotype. XPMC2 of xenopus rescues several different yeast mitotic catastrophe mutants defective in Weel/Mikl kinase function. The XPMC2 protein is localised in the nucleus in Xenopus oocytes. The new protein is the human orthologue of this gene.

The new protein can find application in modulating/blocking the cell cycle.

strong similarity to XPMC2 protein

complete cDMA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="9q34"

Insert length: 2380 bp
Poly A stretch at pos. 2341, polyadenylation signal at pos. 2318

```
   1 AGCGTGCGTG CTGAGGTATG CGCAACGCGT GCGGGGTCTC TTCCGGAGTC
  51 TTTTCCTGGA CGGGGTCCCT GCGGTGGGTG TGTTTCGGCC TGGCCTGGGC
 101 AGGCGCTTGT GCTGCCAGGG CGCCGGGCCC GGGGAGGCCG GGGTCTCGGG
 151 TGGCCGCCGG CCCAGGCGCT GGACGGCAGC AGGATGGGGA AGGCGAAGGT
 201 CCCCGCCTCC AAGCGCGCCC CGAGCAGCCC CGTGGCTAAG CCGGGTCCTG
 251 TCAAGACGCT CACTCGGAAG AAAAACAAGA AGAAAAAAAG GTTTTGGAAA
 301 AGCAAGGCGC GGGAAGTAAG CAAGAAGCCA GCAAGCGGCC CCGGTGCTGT
 351 GGTGCGACCT CCAAAGGCAC CAGAAGACTT TTCTCAAAAC TGGAAGGCGC
 401 TGCAAGAGTG GCTGCTGAAA CAAAAATCTC AGGCCCCAGA AAAGCCTCTT
 451 GTCATCTCTC AGATGGGTTC CAAAAAGAAG CCCAAAATTA TCCAGCAAAA
 501 CAAAAAAGAG ACCTCGCCTC AAGTGAAGGG AGAGGAGATG CCGGCAGGAA
 551 AAGACCAGGA GGCCAGCAGG GGCTCTGTTC CTTCAGGTTC CAAGATGGAC
 601 AGGAGGGCGC CAGTACCTCG CACCAAGGCC AGTGGAACAG AGCACAATAA
 651 GAAAGGAACC AAGGAAAGGA CAAATGGTGA TATTGTTCCA GAACGAGGGG
 701 ACATCGAGCA TAAGAAGCGG AAAGCTAAGG AGGCAGCCCC AGCCCCACCC
 751 ACCGAGGAAG ACATCTGGTT TGACGACGTG GACCCAGCGG ATATCGAAGC
 801 TGCCATAGGT CCAGAGGCGG CCAAGATAGC GAGGAAACAG TTGGGTCAGA
 851 GCGAGGGCAG CGTCAGCCTC AGCCTCGTGA AAGAGCAGGC CTTCGGCGGC
 901 CTGACAAGAG CCTTAGCCTT GGACTGTGAG ATGGTGGGCG TGGGCCCTAA
 951 GGGGGAGGAG AGCATGGCCG CCCGTGTGTC CATCGTGAAC CAGTATGGGA
1001 AGTGCGTTTA TGACAAGTAC GTCAAACCAA CTGAGCCCGT GACGGACTAT
1051 AGGACAGCGG TCAGTGGGAT TCGGCCTGAG AACCTCAAGC AGGGAGAAGA
1101 GCTTGAAGTT GTTCAGAAGG AAGTGGCAGA GATGCTGAAG GGCAGAATTC
1151 TAGTGGGGCA CGCTCTGCAT AATGACCTAA AGGTACTATT TCTTGATCAT
1201 CCAAAAAAGA AGATTCGGGA CACACAGAAA TATAAACCTT TCAAGAGTCA
1251 AGTAAAGAGT GGAAGGCCGT CTCTGAGACT ACTTTCAGAG AAGATCCTTG
1301 GGCTCCAGGT CCAGCAGGCG GAGCACTGTT CAATTCAGGA TGCCCAGGCA
1351 GCAATGAGGC TGTACGTCAT GGTGAAGAAG GAGTGGGAGA GCATGGCCCG
1401 AGACAGGCGC CCCCTGCTGA CTGCTCCAGA CCACTGCAGT GACGACGCCT
1451 AGCAGTCCTG CCCTGCTGCT GCTGCCGCCC CGCTACAGAG GCAATGTGAC
1501 CAGTCACAGG GACAGATCAC ATCTCCCCAG AGTGGCAACT CTGGTGAAAC
1551 CTTTTCAGAA TCATGGCAGA GGGGCGTGGC GTGGTGCTAC TGAGAAGGTC
1601 CTCCTTCCTC TTGACTTTGT GGTCTGAAAC CTGGTCTTAC TGTCCATGTG
1651 TGTTTGGGCC CGGATGGTCA GGGTGGGGAG CAGGGACGGC CATGGGCACG
1701 CCTGGCCACG CTTTACCGAC TGCTGACCCC CTGGGCCAGG TGAGGTTGGG
1751 GCCTGTGGGC CGCCAGTCCA TACGGTGCTG TCACTGCCCA TCTTCGGTGA
1801 CACCCTGGGG TGAGGTGCTC AGCACCTTCC TCTCGAGGAG CCACATTTTC
1851 CTCCTTTGTG TTAGGGGACA TAACAAGCTC TGCTGGGCTT GAGGGACCCA
1901 GACCAGGTGT CTGCAGTCAG CTCCTGAGAC ACAGCTGGCC GGCACAACAG
1951 GTGTTACATC AGGGGTTTCC TGTGGCCGTT TGAACTTTGA GCATTTATCT
2001 AAATTAAATT GGCCCAGGGT TGGCTGGTGG GTCACCCAGC AGAGGCTTCT
2051 CCCCATAGCA CGAGGATGTG TTGCCTGGGC ACGGTGACTG CGGTTATTCC
2101 TGGAGGTCGG CAGACATGCC AACCTTGGGC TATTTGAGCT GGAGAAGCTA
2151 TGTGATGCTA GCCGGTGGCT TTCTGGGCTA GGCCCCAGTT TGAGGCTCCC
2201 CTGGGAACTA GAGCCAGGAA CAGCCAGTGG CACTGACAAG GGGACGGAGT
2251 CCAAGGCGTT ATTGGGCCAC CTGACAGCTG GACAGAAAAG GGGCAGACAC
2301 ACCGAGGATG CGATTTAAAA TAAATGCAGA TGTTTACTTG GAAAAAAAAA
2351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

Entry HSAC2099 from database EMBL:
*** SEQUENCING IN PROGRESS *** Genomic sequence from Human 9q34; HTGS phase 1, 2 unordered pieces.
Score = 5055, P = 0.0e+00, identities = 1011/1011
8 exons Bp 104219-116190

Medline entries

95157530:
Cloning and expression of a Xenopus gene that prevents mitotic catastrophe in fission yeast.

Peptide information for frame 1

ORF from 184 bp to 1449 bp; peptide length: 422
Category: strong similarity to known protein

```
  1 MGKAKVPASK RAPSSPVAKP GPVKTLTRKK NKKKKRFWKS KAREVSKKPA
 51 SGPGAVVRPP KAPEDFSQNW KALQEWLLKQ KSQAPEKPLV ISQMGSKKKP
101 KIIQQNKKET SPQVKGEEMP AGKDQEASRG SVPSGSKMDR RAPVPRTKAS
151 GTEHNKKGTK ERTNGDIVPE RGDIEHKKRK AKEAAPAPPT EEDIWFDDVD
201 PADIEAAIGP EAAKIARKQL GQSEGSVSLS LVKEQAFGGL TRALALDCEM
251 VGVGPKGEES MAARVSIVNQ YGKCVYDKYV KPTEPVTDYR TAVSGIRPEN
301 LKQGEELEVV QKEVAEMLKG RILVGHALHN DLKVLFLDHP KKKIRDTQKY
351 KPFKSQVKSG RPSLRLLSEK ILGLQVQQAE HCSIQDAQAA MRLYVMVKKE
401 WESMARDRRP LLTAPDHCSD DA
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7p10, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3_7p10, frame 1

Report for DKFZphtes3_7p10.1

```
[LENGTH]     422
[MW]         46671.91
[pI]         9.79
[HOMOL]      PIR:S53818 XPMC2 protein - African clawed frog 7e-96
[FUNCAT]     03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-42
[FUNCAT]     01.03.16 polynucleotide degradation  [S. cerevisiae, YGR276c] 2e-19
[FUNCAT]     05.04 translation (initiation, elongation and termination)  [S. cerevisiae, YGL094c] 7e-13
[FUNCAT]     04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation)  [S. cerevisiae, YGL094c] 7e-13
[FUNCAT]     99 unclassified proteins     [S. cerevisiae, YLR107w] 6e-10
[PROSITE]    RGD     1
[PROSITE]    MYRISTYL      4
[PROSITE]    CAMP_PHOSPHO_SITE     2
[PROSITE]    CK2_PHOSPHO_SITE      6
[PROSITE]    TYR_PHOSPHO_SITE      2
[PROSITE]    GLYCOSAMINOGLYCAN     1
[PROSITE]    PKC_PHOSPHO_SITE      8
[KW]         All_Alpha
[KW]         LOW_COMPLEXITY    11.37 %

SEQ   MGKAKVPASKRAPSSPVAKPGPVKTLTRKKNKKKKRFWKSKAREVSKKPASGPGAVVRPP
SEG   ......................xxxxxxxxxxxxxxxxx....................
PRD   cccccccccccccccccccccccccchhhhhhhhhhhhhhhhccccccccccccccccc

SEQ   KAPEDFSQNWKALQEWLLKQKSQAPEKPLVISQMGSKKKPKIIQQNKKETSPQVKGEEMP
SEG   ..................................xxxxxxxxxxxx............
PRD   ccccccccchhhhhhhhhhhhhhccccccccccccccccccceeeecccccccccccccccee
```

```
SEQ   AGKDQEASRGSVPSGSKMDRRAPVPRTKASGTEHNKKGTKERTNGDIVPERGDIEHKKRK
SEG   ......................................................xxxxxx
PRD   ecccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhh

SEQ   AKEAAPAPPTEEDIWFDDVDPADIEAAIGPEAAKIARKQLGQSEGSVSLSLVKEQAFGGL
SEG   xxxxxxxxxxxx................................................
PRD   hhhhcccccccccceeeeccccchhhhhcchhhhhhhhhcccccchhhhhhhhhhhhhhh

SEQ   TRALALDCEMVGVGPKGEESMAARVSIVNQYGKCVYDKYVKPTEPVTDYRTAVSGIRPEN
SEG   ............................................................
PRD   hhhccccccccccccchhhhhhhhhhcccccccceeeeeecccccccccccccccccccc

SEQ   LKQGEELEVVQKEVAEMLKGRILVGHALHNDLKVLFLDHPKKKIRDTQKYKPFKSQVKSG
SEG   ............................................................
PRD   ccccchhhhhhhhhhhhhhcceeeeccchhhhhhhhcccccccccceeeccccccccccc

SEQ   RPSLRLLSEKILGLQVQQAEHCSIQDAQAAMRLYVMVKKEWESMARDRRPLLTAPDHCSD
SEG   ............................................................
PRD   chhhhhhhhhhhhhhhccccccchhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccc

SEQ   DA
SEG   ..
PRD   cc
```

Prosite for DKFZphtes3_7p10.1

PS00002	51->55	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	107->111	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	156->160	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	96->99	PKC_PHOSPHO_SITE	PDOC00005
PS00005	347->350	PKC_PHOSPHO_SITE	PDOC00005
PS00005	359->362	PKC_PHOSPHO_SITE	PDOC00005
PS00005	363->366	PKC_PHOSPHO_SITE	PDOC00005
PS00005	368->371	PKC_PHOSPHO_SITE	PDOC00005
PS00006	136->140	CK2_PHOSPHO_SITE	PDOC00006
PS00006	150->154	CK2_PHOSPHO_SITE	PDOC00006
PS00006	163->167	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	383->387	CK2_PHOSPHO_SITE	PDOC00006
PS00006	413->417	CK2_PHOSPHO_SITE	PDOC00006
PS00007	343->351	TYR_PHOSPHO_SITE	PDOC00007
PS00007	342->351	TYR_PHOSPHO_SITE	PDOC00007
PS00008	130->136	MYRISTYL	PDOC00008
PS00008	151->157	MYRISTYL	PDOC00008
PS00008	221->227	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008
PS00016	171->174	RGD	PDOC00016

(No Pfam data available for DKFZphtes3_7p10.1)

DKFZphtes3_7p9
----

group: nucleic acid management

DKFZphtes3_7p9 encodes a novel 691 amino acid protein with similarity to human nuclear domain 10 protein NDP52.

The nuclear domain (ND)10 also described as POD or Kr bodies is involved in the development of acute promyelocytic leukemia and virus-host interactions. The NDP52 protein is part of this complex structure. In vivo, NDP52 is transcribed in all human tissues, but is redistributed upon viral infection and interferon treatment. ND10 plays an important role in the viral life cycle.
The novel protein is similar to NDP52. It contains three leucine zippers and a RGD cell attachment site. This protein seems to be a novel part of the ND819) complex.

The new protein can find application in modulation of viral infections and tumour events.

similarity to nuclear domain 10 protein NDP52

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="329.1 cR from top of Chr12 linkage group"

Insert length: 3003 bp
Poly A stretch at pos. 2957, no polyadenylation signal found

```
   1 AAGGTGAGGG GAACAGCTGA TCCGTCTGTT GGGAGGACAG ATATCTCAAG
  51 GCCAGGATGG AAGAATCACC ACTAAGCCGG GCACCATCCC GTGGTGGAGT
 101 CAACTTTCTC AATGTAGCCC GGACCTACAT CCCCAACACC AAGGTGGAAT
 151 GTCACTACAC CCTTCCCCCA GGCACCATGC CCAGTGCCAG TGACTGGATT
 201 GGCATCTTCA AGGTGGAGGC TGCCTGTGTT CGGGATTACC ACACATTTGT
 251 GTGGTCTTCC GTGCCTGAAA GTACAACTGA TGGTTCCCCC ATTCACACCA
 301 GTGTCCAGTT CCAAGCCAGC TACCTGCCCA AACCAGGAGC TCAGCTCTAC
 351 CAGTTCCGAT ATGTGAACCG CCAGGGCCAG GTGTGTGGGC AGAGCCCCCC
 401 TTTCCAGTTC CGAGAGCCAA GGCCCATGGA TGAACTGGTG ACCCTGGAGG
 451 AGGCTGATGG GGGCTCTGAC ATCCTGCTGG TTGTCCCCAA GGCAACTGTG
 501 TTACAGAACC AGCTCGATGA GAGCCAGCAA GAACGGAATG ACCTGATGCA
 551 GCTGAAGCTA CAGCTGGAGG GACAGGTGAC AGAGCTGAGG AGCCGAGTGC
 601 AGGAGCTCGA GAGGGCTCTG GCAACTGCCA GGCAGGAGCA CACGGAGCTG
 651 ATGGAACAGT ACAAGGGGAT TTCCCGGTCC CATGGGGAGA TCACAGAAGA
 701 GAGGGACATC CTGAGCCGGC AACAGGGAGA CCATGTGGCA CGCATCCTGG
 751 AGCTAGAGGA TGACATCCAG ACCATCAGTG AGAAAGTGCT GACGAAGGAA
 801 GTGGAGCTGG ACAGGCTTAG AGACACAGTG AAGGCCCTGA CTCGGGAACA
 851 AGAGAAGCTC CTTGGGCAAC TGAAAGAAGT ACAAGCAGAC AAGGAGCAAA
 901 GTGAGGCTGA GCTCCAAGTG GCACAACAGG AGAACCATCA CTTAAATTTG
 951 GACCTGAAGG AGGCGAAGAG CTGGCAAGAG GAGCAGAGTG CTCAGGCTCA
1001 GCGACTGAAA GACAAGGTGG CCCAGATGAA GGACACCCTA GGCCAGGCCC
1051 AGCAGCGGGT GGCCGAGCTG GAGCCCTTGA AGGAGCAGCT TCGAGGGGCC
1101 CAGGAGCTTG CAGCCTCAAG CCAGCAGAAA GCCACCCTTC TTGGGGAGGA
1151 GTTGGCCAGC GCAGCAGCAG CCAGGGACCG CACCATAGCC GAACTACACC
1201 GCAGCCGCCT GGAAGTGGCT GAAGTTAACG GCAGGCTGGC TGAGCTCGGT
1251 TTGCACTTGA AGGAAGAAAA ATGCCAATGG AGCAAGGAGC GGGCAGGGCT
1301 GCTGCAGAGT GTGGAGGCAG AGAAGGACAA GATCCTGAAG CTGAGTGCAG
1351 AGATACTTCG ATTGGAGAAG GCAGTTCAGG AGGAGAGGAC CCAAAACCAA
1401 GTGTTCAAGA CTGAGCTGGC CCGGGAGAAG GATTCTAGCC TGGTACAGTT
1451 GTCAGAAAGT AAGCGGGAGC TGACAGAGCT GCGGTCAGCC CTGCGTGTGC
1501 TCCAGAAGGA AAAGGAGCAG TTACAGGAGG AGAAACAGGA ATTGCTAGAG
1551 TACATGAGAA AGCTAGAGGC CCGCCTGGAG AAGGTGGCAG ATGAGAAGTG
1601 GAATGAGGAT GCCACCACAG AGGATGAGGA GGCCGCTGTG GGGCTGAGCT
1651 GCCCGGCAGC TCTGACAGAC TCAGAGGACG AGTCCCCAGA AGACATGAGG
1701 CTCCCACCCT ATGGCCTTTG TGAGCGTGGA GACCCAGGCT CCTCTCCTGC
1751 TGGGCCTCGA GAGGCTTCTC CCCTTGTTGT CATCAGCCAG CCGGCTCCCA
1801 TTTCTCCTCA CCTCTCTGGG CCAGCTGAGG ACAGTAGCTC TGACTCGGAG
1851 GCTGAAGATG AGAAGTCAGT CCTGATGGCA GCTGTGCAGA GTGGGGGTGA
1901 GGAGGCCAAC TTACTGCTTC CTGAACTGGG CAGTGCCTTC TATGACATGG
1951 CCAGTGGCTT TACAGTGGGT ACCCTGTCAG AAACCAGCAC TGGGGGCCCT
2001 GCCACCCCCA CATGGAAGGA GTGTCCTATC TGTAAGGAGC GCTTTCCTGC
2051 TGAGAGTGAC AAGGATGCCC TGGAGGACCA CATGGATGGA CACTTCTTTT
2101 TCAGCACCCA GGACCCCTTC ACCTTTGAGT GATCTTACTC CCTCGTACAT
2151 GCACAAATAC ACACTCATGC ACACACACAC TCACACACAT GCATACACTT
2201 AGGTTTCATG CCCATTTTCT ATCACACTGG GCTCCATGAT ATTCTGTTCC
2251 CTAAGAACTG CTTCTGTGTG CCCTGTTTTC ATCCCAAGAT TTCTCACTTC
2301 ATCCTCTCCT ACCTGGCTCT TTTGTCCCAG GGAGGGGTCC TGTTCGGAAG
2351 CAGTGGCTGA ATTTATCCCC TGAAAGTGGT TTTGGAGGAA CCGGGATGGA
2401 GGAGGCCTTC CCCTGTGGGA ATAGAATCGT CCACTCCTAG CCCTGGTTGC
```

```
2451 TTCTGATACA CAGCCACTGC ACACACACAC TCACACTCAC ACTCCCTTGT
2501 CTGATGCCCC AAAGCCAATT CCTGGGGCAC CCTACCCTCT CTTATTTGGA
2551 GTTTCCGTTG GTTTACCTGA GTTTTCTCTG GGGTCTGCAC AGAGGCAGCA
2601 GCATGGACAT CATGGCCTCT CAGGTCCCTT TTGGTTCTCA GTTTCATTGG
2651 TTCCTCTTTC TGTTCCCCCA TTGACTTCTG TGCCCCACCC TAGCCTTTTC
2701 CATAACCTTA GGTATTCAGT TTGGAGGGGT TTTTTGTATT TTTGAGGATT
2751 CCTGTATTCT GTATCCTCTC CTCGCATCTC CTCACATGGA AAGAAATAAT
2801 GTATTTGTGC CTTCTGTGAG GAATGGGGGG AACAAGTGGT CCCAGGTATC
2851 CCCATTTCCA AGGCCCCCCT CCCTCTCCAG GTCCCCCCAC AGCAATAAAA
2901 GCTTCCCCCT GATATCCATC CCTTTGTAGT TTGAACAAAT ATATTTATAT
2951 GATATGTAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3001 AAA
```

BLAST Results

Entry HS189353 from database EMBL:
human STS WI-11261.
Score = 2191, P = 1.4e-92, identities = 463/485

Medline entries

95310349:
Molecular characterization of NDP52, a novel protein of the nuclear domain 10, which is redistributed upon virus infection and interferon treatment.

97375672:
Cellular localization, expression, and structure of the nuclear dot protein 52.

Peptide information for frame 3

ORF from 57 bp to 2129 bp; peptide length: 691
Category: similarity to known protein
Prosite motifs: RGD (557-560)
LEUCINE_ZIPPER (163-185)
LEUCINE_ZIPPER (475-497)
LEUCINE_ZIPPER (482-504)

```
  1 MEESPLSRAP SRGGVNFLNV ARTYIPNTKV ECHYTLPPGT MPSASDWIGI
 51 FKVEAACVRD YHTFVWSSVP ESTTDGSPIH TSVQFQASYL PKPGAQLYQF
101 RYVNRQGQVC GQSPPFQFRE PRPMDELVTL EEADGGSDIL LVVPKATVLQ
151 NQLDESQQER NDLMQLKLQL EGQVTELRSR VQELERALAT ARQEHTELME
201 QYKGISRSHG EITEERDILS RQQGDHVARI LELEDDIQTI SEKVLTKEVE
251 LDRLRDTVKA LTREQEKLLG QLKEVQADKE QSEAELQVAQ QENHHLNLDL
301 KEAKSWQEEQ SAQAQRLKDK VAQMKDTLGQ AQQRVAELEP LKEQLRGAQE
351 LAASSQQKAT LLGEELASAA AARDRTIAEL HRSRLEVAEV NGRLAELGLH
401 LKEEKCQWSK ERAGLLQSVE AEKDKILKLS AEILRLEKAV QEERTQNQVF
451 KTELAREKDS SLVQLSESKR ELTELRSALR VLQKEKEQLQ EEKQELLEYM
501 RKLEARLEKV ADEKWNEDAT TEDEEAAVGL SCPAALTDSE DESPEDMRLP
551 PYGLCERGDP GSSPAGPREA SPLVVISQPA PISPHLSGPA EDSSSDSEAE
601 DEKSVLMAAV QSGGEEANLL LPELGSAFYD MASGFTVGTL SETSTGGPAT
651 PTWKECPICK ERFPAESDKD ALEDHMDGHF FFSTQDPFTF E
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_7p9, frame 3

PIR:A56733 nuclear domain 10 protein NDP52 - human, N = 2, Score = 307, P = 7.7e-28

TREMBL:AB008852_1 gene: "NDP"; product: "NDP52"; Bos taurus mRNA for NDP52, complete cds., N = 2, Score = 302, P = 4e-27

TREMBL:AC004549_1 gene: "WUGSC:H_RG459N13.1"; product: "TXBP151"; Homo sapiens BAC clone RG459N13 from 7p15, complete sequence., N = 2, Score = 275, P = 2.3e-25

```
PIR:G02043 TXBP151 - human, N = 2, Score = 270, P = 8.5e-25

TREMBL:DM35816_4 gene: "zip"; product: "nonmuscle myosin-II heavy
chain"; Drosophila melanogaster nonmuscle myosin-II heavy chain (zip)
gene, complete cds., N = 1, Score = 254, P = 1.4e-17


>PIR:A56733 nuclear domain 10 protein NDP52 - human
          Length = 446

  HSPs:

 Score = 307 (46.1 bits), Expect = 7.7e-28, Sum P(2) = 7.7e-28
 Identities = 104/323 (32%), Positives = 158/323 (48%)

Query:    15 VNFLNVARTYIPNTKVECHYTLPPGTMPSASDWIGIFKVEAACVRDYHTFVWSSVPESTT 74
             V F +V + YIP   V CHYT       +P   DWIGIF+V      R+Y+TF+W ++P
Sbjct:    23 VIFNSVEKFYIPGGDVTCHYTFTQHFIPRRKDWIGIFRVGWKTTREYYTFMWVTLPIDLN 82

Query:    75 DGSPIHTSVQFQASYLPKPGAQLYQFRYVNRQGQVCGQSPPFQFREPRPMDELVTLEEAD 134
             + S      VQF+A YLPK    + YQF YV+  G V G S PFQFR       D LV    +
Sbjct:    83 NKSAKQQEVQFKAYYLPKDD-EYYQFCYVDEDGVVRGASIPFQFRPENEEDILVVTTQ-- 139

Query:   135 GGSDILLVVPKATVLQNQ-LDES---QQERNDLMQLKLQLEGQVTE-LRSRVQELERALA 189
             G  + +     K     +NQ L +S     Q++N  MQ +LQ + +  E L+S  ++LE  +
Sbjct:   140 GEVEEIEQHNKELCKENQELKDSCISLQKQNSDMQAELQKKQEELETLQSINKKLELKVK 199

Query:   190 TARQE-HTELMEQYKGISRSHGEITEERDI-LSRQQGDHVARILELEDDIQTISEKVLTK 247
                +     TEL+ Q K  ++         E+  I + + Q       +  E+E  +Q   +K  T+
Sbjct:   200 EQKDYWETELL-QLKEQNQKMSSENEKMGIRVDQLQAQLSTQEKEMEKLVQGDQDK--TE 256

Query:   248 EVE-LDRLRDTVKALTREQEKLLGQLKEVQADKEQSEAELQVAQQENHHLNLDLKEAKSW 306
             ++E L +   D +       EQ K    +L++      +Q+E         QQE     N DL  +  S
Sbjct:   257 QLEQLKKENDHLFLSLTEQRKDQKKLEQTVEQMKQNETTAMKKQQELMDENFDLSKRLSE 316

Query:   307 QEEQSAQAQRLKDKVAQMKDTLGQAQQRV 335
              E       QR K+++      D L +    R+
Sbjct:   317 NEIICNALQRQKERLEGENDLLKRENSRL 345

 Score = 304 (45.6 bits), Expect = 2.1e-27, Sum P(2) = 2.1e-27
 Identities = 98/337 (29%), Positives = 163/337 (48%)

Query:    15 VNFLNVARTYIPNTKVECHYTLPPGTMPSASDWIGIFKVEAACVRDYHTFVWSSVPESTT 74
             V F +V + YIP   V CHYT       +P   DWIGIF+V      R+Y+TF+W ++P
Sbjct:    23 VIFNSVEKFYIPGGDVTCHYTFTQHFIPRRKDWIGIFRVGWKTTREYYTFMWVTLPIDLN 82

Query:    75 DGSPIHTSVQFQASYLPKPGAQLYQFRYVNRQGQVCGQSPPFQFREPRPMDELVTLEEAD 134
             + S      VQF+A YLPK    + YQF YV+  G V G S PFQFR    P +E
Sbjct:    83 NKSAKQQEVQFKAYYLPKDD-EYYQFCYVDEDGVVRGASIPFQFR---PENE-------- 130

Query:   135 GGSDILLVVPKATVLQNQLDESQQERNDLMQLKLQLEGQVTELRSRVQELERALATARQE 194
               DIL+V           Q +++E +Q     +L +     +L+       L+ +   +++   L     +QE
Sbjct:   131 --EDILVVTT-----QGEVEEIEQHNKELCKENQELKDSCISLQKQNSDMQAELQK-KQE 182

Query:   195 HTELMEQYKGISRSHGEITEERDILSRQQGDH-VARILELEDDIQTISEKVLTKEVELDR 253
              E ++               I ++ ++  ++Q D+     +L+L++  Q +S +      + +D+
Sbjct:   183 ELETLQS----------INKKLELKVKEQKDYWETELLQLKEQNQKMSSENEKMGIRVDQ 232

Query:   254 LRDTVKALTREQEKLL--GQLKEVQAD---KEQSEAELQVAQQENHHLNLDLKEAKSWQE 308
             L+  +     +E EKL+     Q K    Q +      KE         L + +Q            L+       + Q
Sbjct:   233 LQAQLSTQEKEMEKLVQGDQDKTEQLEQLKKENDHLFLSLTEQRKDQKKLEQTVEQMKQN 292

Query:   309 EQSA--QAQRLKDKVAQMKDTLGQAQQRVAELEPLKEQLRGAQEL 351
             E +A    +  Q L D+      +     L + +      L+   KE+L G  +L
Sbjct:   293 ETTAMKKQQELMDENFDLSKRLSENEIICNALQRQKERLEGENDL 337

 Score = 124 (18.6 bits), Expect = 2.3e-06, Sum P(2) = 2.3e-06
 Identities = 53/227 (23%), Positives = 113/227 (49%)

Query:   138 DILLVVPKATVLQNQLDESQQERNDLMQLKLQLEGQVTELRSRVQELERALATARQEHTE 197
             DIL+V           Q +++E +Q     +L +     +L+       L+ +   +++   L     +QE  E
Sbjct:   132 DILVVTT-----QGEVEEIEQHNKELCKENQELKDSCISLQKQNSDMQAELQK-KQEELE 185

Query:   198 LMEQYKGISRSHGEITEERDILSRQQGDH-VARILELEDDIQTISEKVLTKEVELDRLRD 256
              ++               I ++ ++  ++Q D+     +L+L++  Q +S +       + +D+L+
Sbjct:   186 TLQS----------INKKLELKVKEQKDYWETELLQLKEQNQKMSSENEKMGIRVDQLQA 235

Query:   257 TVKALTREQEKLLGQLKEVQADKEQSEAELQVAQQENHHLNLDLKEAKSWQEEQSAQAQR 316
                +     +E EKL           VQ D++++E +L+     ++EN HL L L E +     Q++          ++
Sbjct:   236 QLSTQEKEMEKL------VQGDQDKTE-QLEQLKKENDHLFLSLTEQRKDQKKLEQTVEQ 288
```

```
Query:   317 LK-DKVAQMKDTLGQAQQRVAELEPLKEQLRGAQELA-ASSQQKATLLGE 364
             +K ++   MK     + Q+ + E   L ++L    + +  A  +QK  L GE
Sbjct:   289 MKQNETTAMK----KQQELMDENFDLSKRLSENEIICNALQRQKERLEGE 334

 Score = 103 (15.5 bits), Expect = 4.4e-04, Sum P(2) = 4.4e-04
 Identities = 63/278 (22%), Positives = 123/278 (44%)

Query:   299 DLKEAKSWQEEQSAQAQRLKDKVAQMK----DTLGQAQQRVAELEPLKEQLRGAQELAAS 354
             +++E +    +E    + Q LKD    ++      D    + Q++  ELE L + +    EL
Sbjct:   141 EVEEIEQHNKELCKENQELKDSCISLQKQNSDMQAELQKKQEELETL-QSINKKLELKVK 199

Query:   355 SQQKATLLGEELASAAAARDRTIAELHRSRLEVAEVNGRLAELGLHLKEEKCQWSKERAG 414
              Q+          EL           +  +E  +   + V ++  +L+         + E+  Q  +++
Sbjct:   200 EQKD--YWETELLQLKEQNQKMSSENEKMGIRVDQLQAQLSTQEKEM-EKLVQGDQDKTE 256

Query:   415 LLQSVEAEKDKI-LKLSAEIL---RLEKAVQEERTQNQVFKTELAREKDSSLVQLSESKR 470
              L+ ++ E D + L L+ +      +LE+ V E+  QN+   T + ++++          SKR
Sbjct:   257 QLEQLKKENDHLFLSLTEQRKDQKKLEQTV-EQMKQNET--TAMKKQQELMDENFDLSKR 313

Query:   471 ELTELRSALRVLQKEKEQLQEEKQELLEYMRKLEARLEKVADEKWNE---DATTEDEEAA 527
              L+E           LQ++KE+L+ E  +LL    ++  +RL          +N        T DE  A
Sbjct:   314 -LSENEIICNALQRQKERLEGEN-DLL---KRENSRLLSYMGLDFNSLPYQVPTSDEGGA 368

Query:   528 ---VGLSCPAALTD-SEDESPEDMRLPPYGLCERGDPGSSPAGPREASPL 573
                  GL+        +       E  SP  + +      +C+  D              ++  PL
Sbjct:   369 RQNPGLAYGNPYSGIQESSSPSPLSIKKCPICKADDICDHTLEQQQMQPL 418

 Score = 64 (9.6 bits), Expect = 7.7e-28, Sum P(2) = 7.7e-28
 Identities = 13/29 (44%), Positives = 17/29 (58%)

Query:   651 PTWKECPICKERFPAESDKDALEDHMDGH 679
             P     CPIC + FPA ++K    EDH+  H
Sbjct:   417 PLCFNCPICDKIFPA-TEKQIFEDHVFCH 444

 Score = 64 (9.6 bits), Expect = 5.8e+00, Sum P(2) = 1.0e+00
 Identities = 26/90 (28%), Positives = 45/90 (50%)

Query:   470 RELTELRSALRVLQKEKEQLQEE---KQELLEYMRKLEARLE-KVADEK--W-------- 515
             +E  EL+ +     LQK+     +Q E     KQE LE ++ +   +LE KV ++K   W
Sbjct:   154 KENQELKDSCISLQKQNSDMQAELQKKQEELETLQSINKKLELKVKEQKDYWETELLQLK 213

Query:   516 --NEDATTEDEEAAVGLS-CPAALTDSEDE 542
                 N+   ++E+E+   +  +        A L+  E E
Sbjct:   214 EQNQKMSSENEKMGIRVDQLQAQLSTQEKE 243

 Score = 47 (7.1 bits), Expect = 4.6e-26, Sum P(2) = 4.6e-26
 Identities = 11/30 (36%), Positives = 17/30 (56%)

Query:   631 MASGFTVGTLSETSTGGPATPTWKECPICK 660
             +A G      + E+S++   P +     K+CPICK
Sbjct:   374 LAYGNPYSGIQESSSPSPLSI--KKCPICK 401
```

Pedant information for DKFZphtes3_7p9, frame 3

Report for DKFZphtes3_7p9.3

```
[LENGTH]    691
[MW]        77336.52
[pI]        4.77
[HOMOL]     PIR:A56733 nuclear domain 10 protein NDP52 - human 2e-29
[FUNCAT]    09.10 nuclear biogenesis     [S. cerevisiae, YDR356w] 2e-11
[FUNCAT]    30.04 organization of cytoskeleton    [S. cerevisiae, YDR356w] 2e-11
[FUNCAT]    08.07 vesicular transport (golgi network, etc.)     [S. cerevisiae, YDL058w]
2e-11
[FUNCAT]    03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 2e-11
[FUNCAT]    30.03 organization of cytoplasm      [S. cerevisiae, YDL058w] 2e-11
[FUNCAT]    99 unclassified proteins      [S. cerevisiae, YLR309c] 2e-08
[FUNCAT]    03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w
MYO1 - myosin-1 isoform] 3e-07
[FUNCAT]    08.22 cytoskeleton-dependent transport       [S. cerevisiae, YHR023w MYO1 -
myosin-1 isoform] 3e-07
[FUNCAT]    03.25 cytokinesis     [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-07
[FUNCAT]    09.13 biogenesis of chromosome structure      [S. cerevisiae, YJL074c] 4e-07
[FUNCAT]    30.10 nuclear organization     [S. cerevisiae, YNL250w] 4e-06
[FUNCAT]    03.07 pheromone response, mating-type determination, sex-specific proteins
      [S. cerevisiae, YBR289w] 4e-06
```

[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YBR289w] 4e-06
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YBR289w] 4e-06
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 4e-06
[FUNCAT] 03.13 meiosis [S. cerevisiae, YNL250w] 4e-06
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1643] 1e-05
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 4e-05
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 4e-05
[FUNCAT] 08.19 cellular import [S. cerevisiae, YNL243w] 7e-05
[FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YNL243w] 7e-05
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YNL243w] 7e-05
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c] 2e-04
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 2e-04
[BLOCKS] BL00682B ZP domain proteins
[EC] 3.6.1.32 Myosin ATPase 1e-13
[PIRKW] nucleus 6e-10
[PIRKW] phosphotransferase 2e-07
[PIRKW] duplication 9e-07
[PIRKW] citrulline 1e-09
[PIRKW] tandem repeat 1e-13
[PIRKW] heart 5e-11
[PIRKW] endocytosis 5e-09
[PIRKW] polymorphism 3e-06
[PIRKW] cornified cell envelope 1e-06
[PIRKW] transmembrane protein 6e-12
[PIRKW] serine/threonine-specific protein kinase 2e-07
[PIRKW] cell wall 1e-06
[PIRKW] zinc finger 5e-09
[PIRKW] metal binding 5e-09
[PIRKW] DNA binding 8e-08
[PIRKW] muscle contraction 1e-11
[PIRKW] IgG constant region-binding 1e-06
[PIRKW] acetylated amino end 4e-09
[PIRKW] actin binding 1e-13
[PIRKW] mitosis 9e-09
[PIRKW] microtubule binding 9e-09
[PIRKW] ATP 1e-13
[PIRKW] thick filament 1e-10
[PIRKW] phosphoprotein 1e-13
[PIRKW] epidermis 1e-06
[PIRKW] leucine zipper 1e-07
[PIRKW] glycoprotein 4e-07
[PIRKW] skeletal muscle 4e-10
[PIRKW] disulfide bond 1e-07
[PIRKW] calcium binding 1e-09
[PIRKW] alternative splicing 1e-10
[PIRKW] coiled coil 1e-13
[PIRKW] P-loop 1e-13
[PIRKW] heptad repeat 6e-10
[PIRKW] methylated amino acid 1e-13
[PIRKW] basement membrane 3e-06
[PIRKW] immunoglobulin receptor 2e-07
[PIRKW] peripheral membrane protein 5e-09
[PIRKW] dimer 1e-07
[PIRKW] cardiac muscle 1e-10
[PIRKW] extracellular matrix 3e-06
[PIRKW] hydrolase 1e-13
[PIRKW] microtubule 6e-10
[PIRKW] muscle 2e-09
[PIRKW] membrane protein 3e-06
[PIRKW] EF hand 1e-09
[PIRKW] cytoskeleton 6e-12
[PIRKW] hair 1e-09
[PIRKW] calmodulin binding 5e-09
[PIRKW] Golgi apparatus 3e-08
[SUPFAM] myosin heavy chain 1e-13
[SUPFAM] conserved hypothetical P115 protein 1e-08
[SUPFAM] hypothetical protein YJL074c 5e-07
[SUPFAM] centromere protein E 9e-09
[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 2e-07
[SUPFAM] calmodulin repeat homology 1e-09
[SUPFAM] myosin motor domain homology 1e-13
[SUPFAM] alpha-actinin actin-binding domain homology 3e-13
[SUPFAM] tropomyosin 3e-07
[SUPFAM] plectin 3e-13
[SUPFAM] trichohyalin 1e-09
[SUPFAM] pleckstrin repeat homology 4e-06
[SUPFAM] ribosomal protein S10 homology 3e-13

```
[SUPFAM]        giantin 3e-08
[SUPFAM]        protein kinase homology 2e-07
[SUPFAM]        protein kinase C zinc-binding repeat homology 4e-06
[SUPFAM]        involucrin 1e-06
[SUPFAM]        kinesin motor domain homology 9e-09
[SUPFAM]        human early endosome antigen 1 5e-09
[SUPFAM]        unassigned kinesin-related proteins 8e-08
[SUPFAM]        M5 protein 3e-08
[SUPFAM]        cytoskeletal keratin 3e-08
[PROSITE]       LEUCINE_ZIPPER 3
[PROSITE]       RGD     1
[PROSITE]       MYRISTYL        6
[PROSITE]       CK2_PHOSPHO_SITE        25
[PROSITE]       PKC_PHOSPHO_SITE        6
[KW]            All_Alpha
[KW]            LOW_COMPLEXITY     9.12 %
[KW]            COILED_COIL       39.36 %

SEQ     MEESPLSRAPSRGGVNFLNVARTYIPNTKVECHYTLPPGTMPSASDWIGIFKVEAACVRD
SEG     ............................................................
PRD     ccccccccccccceeeecceeeeeccccaeeeeecccccccccccceeeeeeeeeeecccc
COILS   ............................................................

SEQ     YHTFVWSSVPESTTDGSPIHTSVQFQASYLPKPGAQLYQFRYVNRQGQVCGQSPPFQFRE
SEG     ............................................................
PRD     eeeeeeeecccccccccchhhhhhhhhhhccccccceeeeeccccccccccccccccccc
COILS   ............................................................

SEQ     PRPMDELVTLEEADGGSDILLVVPKATVLQNQLDESQQERNDLMQLKLQLEGQVTELRSR
SEG     ............................................................
PRD     cccccceeehhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   .........................CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ     VQELERALATARQEHTELMEQYKGISRSHGEITEERDILSRQQGDHVARILELEDDIQTI
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCC.................................

SEQ     SEKVLTKEVELDRLRDTVKALTREQEKLLGQLKEVQADKEQSEAELQVAQQENHHLNLDL
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   .......CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ     KEAKSWQEEQSAQAQRLKDKVAQMKDTLGQAQQRVAELEPLKEQLRGAQELAASSQQKAT
SEG     ..........................................................xx
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCC..CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.CCCCCCCCCCCCCCCCCCC

SEQ     LLGEELASAAAARDRTIAELHRSRLEVAEVNGRLAELGLHLKEEKCQWSKERAGLLQSVE
SEG     xxxxxxxxxxxxxxxxxxxx........................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCC..........................................CCCCCCCCCC

SEQ     AEKDKILKLSAEILRLEKAVQEERTQNQVFKTELAREKDSSLVQLSESKRELTELRSALR
SEG     ............................................................
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCC.............CCCCCCCCCCCCCCCCCCCC

SEQ     VLQKEKEQLQEEKQELLEYMRKLEARLEKVADEKWNEDATTEDEEAAVGLSCPAALTDSE
SEG     .xxxxxxxxxxxxxxxxx..................xxxxxxxxxxx.............
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC..........................

SEQ     DESPEDMRLPPYGLCERGDPGSSPAGPREASPLVVISQPAPISPHLSGPAEDSSSDSEAE
SEG     .................................................xxxxxxxxxxx
PRD     hhhhcccccccccccccccccccccccccccccceeeeeeccccccccccccccccccchh
COILS   ............................................................

SEQ     DEKSVLMAAVQSGGEEANLLLPELGSAFYDMASGFTVGTLSETSTGGPATPTWKECPICK
SEG     xx..........................................................
PRD     hhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccccccc
COILS   ............................................................

SEQ     ERFPAESDKDALEDHMDGHFFFSTQDPFTFE
SEG     ...............................
PRD     cccccccchhhhhhccccceeeccccccccc
COILS   ...............................
```

Prosite for DKFZphtes3_7p9.3

PS00005	190->193	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	468->471	PKC_PHOSPHO_SITE	PDOC00005
PS00005	652->655	PKC_PHOSPHO_SITE	PDOC00005
PS00005	667->670	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	43->47	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	239->243	CK2_PHOSPHO_SITE	PDOC00006
PS00006	282->286	CK2_PHOSPHO_SITE	PDOC00006
PS00006	305->309	CK2_PHOSPHO_SITE	PDOC00006
PS00006	376->380	CK2_PHOSPHO_SITE	PDOC00006
PS00006	383->387	CK2_PHOSPHO_SITE	PDOC00006
PS00006	468->472	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	537->541	CK2_PHOSPHO_SITE	PDOC00006
PS00006	539->543	CK2_PHOSPHO_SITE	PDOC00006
PS00006	543->547	CK2_PHOSPHO_SITE	PDOC00006
PS00006	593->597	CK2_PHOSPHO_SITE	PDOC00006
PS00006	595->599	CK2_PHOSPHO_SITE	PDOC00006
PS00006	597->601	CK2_PHOSPHO_SITE	PDOC00006
PS00006	612->616	CK2_PHOSPHO_SITE	PDOC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDOC00006
PS00006	652->656	CK2_PHOSPHO_SITE	PDOC00006
PS00006	667->671	CK2_PHOSPHO_SITE	PDOC00006
PS00006	683->687	CK2_PHOSPHO_SITE	PDOC00006
PS00008	39->45	MYRISTYL	PDOC00008
PS00008	107->113	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	414->420	MYRISTYL	PDOC00008
PS00008	561->567	MYRISTYL	PDOC00008
PS00008	613->619	MYRISTYL	PDOC00008
PS00016	557->560	RGD	PDOC00016
PS00029	163->185	LEUCINE_ZIPPER	PDOC00029
PS00029	475->497	LEUCINE_ZIPPER	PDOC00029
PS00029	482->504	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3_7p9.3)

DKFZphtes3_8e24
----

group: signal transduction

DKFZphtes3_8e24.3 encodes a novel 658 amino acid putative GTP-binding protein, related to yeast YGL099w and mouse MMR1 putative GTP-binding proteins.

GTP-binding proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

strong similarity to guanine nucleotide binding proteins

complete cDNA, complete cds, potential start at Bp 31, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 3290 bp
Poly A stretch at pos. 3269, polyadenylation signal at pos. 3251

```
   1 CGTCCAGCGG TCGTGTTGCC ATGGGCCGGA GGAGAGCCCC GGCCGGTGGG
  51 TCGCTGGGAC GGGCCCTTAT GCGCCATCAG ACTCAGCGGA GCCGAAGCCA
 101 TCGTCACACT GACTCCTGGT TGCACACAAG TGAACTCAAT GATGGCTATG
 151 ATTGGGGTCG TCTTAATCTT CAGTCAGTGA CTGAACAGAG CTCCCTTGAT
 201 GACTTCCTTG CTACTGCAGA ACTTGCAGGA ACAGAGTTTG TAGCTGAAAA
 251 ACTTAATATT AAGTTTGTGC CTGCTGAGGC TAGAACTGGA CTACTGTCTT
 301 TCGAGGAGAG CCAGAGAATT AAGAAGCTCC ATGAAGAAAA CAAACAGTTC
 351 TTGTGTATAC CGAGGAGACC AAACTGGAAC CAAAATACTA CCCCAGAAGA
 401 ACTCAAACAA GCAGAGAAAG ATAACTTTCT AGAATGGAGA CGTCAGCTTG
 451 TCCGGCTAGA AGAGGAACAG AAGCTGATAT TGACTCCATT TGAACGAAAT
 501 TTGGACTTTT GGCGCCAGCT CTGGAGAGTC ATTGAGAGAA GTGATATTGT
 551 GGTCCAGATA GTAGATGCTC GAAACCCACT CCTGTTTAGA TGTGAGGATT
 601 TGGAATGTTA TGTGAAAGAA ATGGATGCCA ATAAGGAGAA CGTCATTCTG
 651 ATCAACAAGG CAGACTTGCT GACTGCTGAG CAGCGGAGTG CCTGGGCCAT
 701 GTACTTCGAA AAAGAAGATG TGAAGGTTAT TTTCTGGTCA GCTTTGGCCG
 751 GAGCCATTCC CCTGAATGGT GACTCTGAGG AAGAGGCAAA CAGAGATGAT
 801 AGACAAAGCA ACACAACTGA GTTTGGACAT TCCAGTTTCG ACCAGGCTGA
 851 AATTTCCCAC AGTGAATCCG AACATCTCCC AGCTAGGGAT TCTCCTTCAC
 901 TTAGTGAAAA TCCCACAACG GATGAAGATG ACAGTGAGTA TGAGGACTGT
 951 CCAGAGGAGG AGGAAGACGA CTGGCAGACG TGCTCAGAAG AAGACGGTCC
1001 CAAGGAAGAG GACTGCAGCC AGGACTGGAA GGAAAGCTCT ACTGCAGATT
1051 CTGAGGCTCG GAGCAGGAAA ACCCCACAGA AGAGGCAGAT ACACAATTTT
1101 AGCCATCTGG TATCCAAGCA GGAGTTACTG GAGCTCTTTA AGGAGCTACA
1151 CACTGGGAGA AAGGTGAAAG ATGGGCAACT TACGGTCGGA CTGGTGGGCT
1201 ACCCTAATGT TGGTAAGAGT TCAACAATCA ACACCATCAT GGGCAACAAG
1251 AAAGTATCTG TGTCTGCCAC ACCTGGTCAC ACAAAGCACT TTCAGACTCT
1301 CTATGTGGAG CCTGGCCTCT GCCTGTGTGA CTGTCCTGGC TTGGTGATGC
1351 CATCTTTTGT GTCTACCAAG GCAGAAATGA CTTGCAGCGG AATCCTCCCA
1401 ATTGATCAGA TGAGAGATCA TGTTCCTCCT GTATCACTAG TTTGCCAGAA
1451 TATTCCAAGA CATGTTTTAG AAGCTACCTA TGGCATTAAC ATCATAACGC
1501 CTAGAGAGGA TGAAGATCCC CACCGACCTC CAACATCGGA AGAACTGTTG
1551 ACAGCTTATG GATACATGCG AGGATTCATG ACAGCGCATG GACAGCCAGA
1601 CCAGCCTCGA TCTGCGCGCT ACATCCTGAA GGACTATGTC AGTGGTAAGC
1651 TGCTGTACTG CCATCCTCCT CCTGGAAGAG ATCCTGTAAC TTTTCAGCAT
1701 CAACACCAGC GACTCCTAGA GAACAAAATG AACAGTGATG AAATAAAAAT
1751 GCAGCTAGGC AGAAATAAAA AAGCAAAGCA GATTGAAAAT ATCGTTGACA
1801 AAACTTTTTT CCATCAAGAG AATGTGAGGG CTTTGACCAA AGGAGTCCAG
1851 GCTGTGATGG GTTACAAGCC CGGGAGTGGT GTAGTGACTG CATCCACTGC
1901 GAGCTCTGAG AACGGGGCGG GGAAGCCCTG GAAAAAACAT GGCAACAGAA
1951 ATAAAAAAGA AAAAAGTCGT AGACTCTACA AGCACCTGGA TATGTGAGGT
2001 TGGGCTGCAA CAGAAATGTC ATCTGCATTG TGCAGATGGA AAAGAGCAGA
2051 AGCTGCCTGT TGCCTGTGGA ACTGTCCCAA GACACTAGCA CTGTAGAACG
2101 GGCCCTGCTC TTGCAGAGCA CGGCTGCACC CAACAGTCTC CATGTCAAGA
2151 CCAAGGGCCT CCTGGAAACA CCAGCTCTGA CAAAAAGGAG TCATCTGGGA
2201 GCCCGAGAAT CCTACTCCTG GCCGGGCACA GTGGCTCACG CACCAACATG
2251 GAGAAACCCC GTCTCTACTA AAAATACAAA AAAATTAGCC AGGCGTGGTG
2301 GCGCGCACCT GTAATCCCAG CTACTCGGGA GGCTGAGGCA GGAGAATCAC
2351 TTGAACCAGG GAGGCAGAGT TTGCAGTGAA TGGAGATTGC GCCGCTGCAC
2401 TCCAGCCTGG GCGACAGAGT GAGACTGCAT CACAAGAAAA AAAATTTGCA
2451 AGGGATGGTT CACGAGACAC ATTTGGGACG AAGGTGAAAG AGAAATTCCC
2501 CATTCTGAGT GTCCTAGTTG GGTTCCTCCG ACTCTAAACA AGGGACTTGG
2551 GTTCAGTTAG TGTACAGCGG GGGCTCACGT CCACTAAGGA ACATGTAGAA
2601 TGTAACCACC GGGTGACAGG GAAGCTGCGG TATTTACTAC CTAGCCCCCA
```

```
2651 TCTTCACTGG TTATTCCACT TATTTAAAAT GTCCAGAATA AGCAAATCTC
2701 CATATAGAGG AAGTAGATTA GTGGTTGCTT CGGGATGGGA GGAATGGGAA
2751 GATTGAGGTC TTTCTTTTGC AGTGATAAAA ATGTCCTAAA ATTGACTGTA
2801 GCGATGGTCA CACAACTCTG AATATGCTTA AGACCATTGA ATTACACACT
2851 TTACGTTGGT GAATTGTATG GTATGTAAAT TATAGTTCAA TAACATAGTT
2901 ACAAAAGATA ATCAAAAGCA TGAAAGCACT ATTGATGTGG TTTGGATCTG
2951 TGTCCTCACC GAGTCTCATG TTGAAATGTA AGCCCCCTGG TGGGAGGCGA
3001 TGGGATTATG GGGCAGAGTC CTCACAAACG GTTTAGCACC ACCCGCTCAG
3051 TGCTGTTCTC CTGATATTGA GTCCTCATCA CATCTGGTTG CTTCAAAGTG
3101 TGTGGTGCCT CCCCTCTGTC TCCCTCCTGC TCTGGCCATA TAAGATGTGC
3151 CTGCTTCTCC TTCGCCTTCT AACATGATTG TAAGTTTCCT GAGGCCTCCC
3201 TAGAAGCAAA AGCTGCTGTG CTTCCTGTAC CATCTACTGG ACCGTGAGCC
3251 AATTAAACCT CTTTTCTTTA TAAAAAAAAA AAAAAAAAGG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 21 bp to 1994 bp; peptide length: 658
Category: strong similarity to known protein

```
  1 MGRRRAPAGG SLGRALMRHQ TQRSRSHRHT DSWLHTSELN DGYDWGRLNL
 51 QSVTEQSSLD DFLATAELAG TEFVAEKLNI KFVPAEARTG LLSFEESQRI
101 KKLHEENKQF LCIPRRPNWN QNTTPEELKQ AEKDNFLEWR RQLVRLEEEQ
151 KLILTPFERN LDFWRQLWRV IERSDIVVQI VDARNPLLFR CEDLECYVKE
201 MDANKENVIL INKADLLTAE QRSAWAMYFE KEDVKVIFWS ALAGAIPLNG
251 DSEEEANRDD RQSNTTEFGH SSFDQAEISH SESEHLPARD SPSLSENPTT
301 DEDDSEYEDC PEEEEDDWQT CSEEDGPKEE DCSQDWKESS TADSEARSRK
351 TPQKRQIHNF SHLVSKQELL ELFKELHTGR KVKDGQLTVG LVGYPNVGKS
401 STINTIMGNK KVSVSATPGH TKHFQTLYVE PGLCLCDCPG LVMPSFVSTK
451 AEMTCSGILP IDQMRDHVPP VSLVCQNIPR HVLEATYGIN IITPREDEDP
501 HRPPTSEELL TAYGYMRGFM TAHGQPDQPR SARYILKDYV SGKLLYCHPP
551 PGRDPVTFQH QHQRLLENKM NSDEIKMQLG RNKKAKQIEN IVDKTFFHQE
601 NVRALTKGVQ AVMGYKPGSG VVTASTASSE NGAGKPWKKH GNRNKKEKSR
651 RLYKHLDM
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8e24, frame 3

SWISSPROT:YAWG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN CHROMOSOME I., N = 3, Score = 560, P = 1.6e-111

PIR:S64106 hypothetical protein YGL099w - yeast (Saccharomyces cerevisiae), N = 2, Score = 544, P = 2.6e-105

TREMBL:CEAF3143_1 gene: "C53H9.2"; Caenorhabditis elegans cosmid C53H9., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:MMR1_MOUSE POSSIBLE GTP-BINDING PROTEIN MMR1., N = 2, Score = 311, P = 7.5e-31

```
>SWISSPROT:YAWG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN
     CHROMOSOME I.
          Length = 616

  HSPs:

 Score = 560 (84.0 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
 Identities = 119/253 (47%), Positives = 163/253 (64%)

Query:    12 LGRALMRHQTQRSRSHRHTDSWLHTSELNDGYDWGRLNLQSVTEQSSLDDFLATAELAGT 71
             LGRA+     T+  R+ +       H  + +        R  L+SVT ++ LD+FL TAEL
```

```
Sbjct:    12 LGRAIQSDFTKNRRNRK--GGLKHIVDSDPKAH--RAALRSVTHETDLDEFLNTAELGEV 67

Query:    72 EFVAEKLNIKFVP-AEARTGLLSFEESQRIKKLHEENKQFLCIPRRPNWNQNTTPEELKQ 130
             EF+AEK N+  +    E     LLS EE+ R K+  E+NK  L IPRRP+W+Q TT  EL +
Sbjct:    68 EFIAEKQNVTVIQNPEQNPFLLSKEEAARSKQKQEKNKDRLTIPRRPHWDQTTTAVELDR 127

Query:   131 AEKDNFLEWRRQLVRLEEEQKLILTPFERNLDFWRQLWRVIERSDIVVQIVDARNPLLFR 190
               E+++FL WRR L +L++ +   I+TPFERNL+ WRQLWRVIERSD+VVQIVDARNPL FR
Sbjct:   128 MERESFLNWRRNLAQLQDVEGFIVTPFERNLEIWRQLWRVIERSDVVVQIVDARNPLFFR 187

Query:   191 CEDLECYVKEMDANKENVILINKADLLTAEQRSAWAMYFEKEDVKVIFWSALAGAIPLNG 250
               LE YVKE+  +K+N +L+NKAD+LT EQR+ W+ YF + ++  +F+SA   A   N
Sbjct:   188 SAHLEQYVKEVGPSKKNFLLVNKADMLTEEQRNYWSSYFNENNIPFLFFSARMAA-EANE 246

Query:   251 DSEEEANRDDRQSN 264
                E+    +  SN
Sbjct:   247 RGEDLETYESTSSN 260
```

Score = 532 (79.8 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
Identities = 131/323 (40%), Positives = 192/323 (59%)

```
Query:   340 STADSEARSRKTPQKRQIHNFSHLVSKQELLELFKELHTGRKVKDGQ--LTVGLVGYPNV 397
             ST+ +E          +  +H+ S + + + L  +F++  +   + DG+  +T GLVGYPNV
Sbjct:   256 STSSNEIPESLQADENDVHS-SRIATLKVLEGIFEKFAS--TLPDGKTKMTFGLVGYPNV 312

Query:   398 GKSSTINTIMGNKKVSVSATPGHTKHFQTLYVEPGLCLCDCPGLVMPSFVSTKAEMTCSG 457
             GKSSTIN ++G+KKVSVS+TPG TKHFQT+ +   + L DCPGLV PSF +T+A++   G
Sbjct:   313 GKSSTINALVGSKKVSVSSTPGKTKHFQTINLSEKVSLLDCPGLVFPSFATTQADLVLDG 372

Query:   458 ILPIDQMRDHVPPVSLVCQNIPRHVLEATYGINI-ITPREDEDPHRPPTSEELLTAYGYM 516
             +LPIDQ+R++  P +L+ + IP+ VLE  Y I I I P E E     P+++E+L  +
Sbjct:   373 VLPIDQLREYTGPSALMAERIPKEVLETLYTIRIRIKPIE-EGGTGVPSAQEVLFPFARS 431

Query:   517 RGFMTAH-GQPDQPRSARYILKDYVSGKLLYCHPPPG--RDPVTFQHQHQRLLENKMNSD 573
             RGFM AH G PD  R+AR +LKDYV+GKLLY HPPP         F  +H + + +  SD
Sbjct:   432 RGFMRAHHGTPDDSRAARILLKDYVNGKLLYVHPPPNYPNSGSEFNKEHHQKIVSA-TSD 490

Query:   574 EIKMQLGR---NKKAKQIEN-IVDKTFFHQEN--VRALTKGVQAVM-G--YKPGSGVVTA 624
              I  +L R   +      E+ +VD  +F QEN  VR + KG   M G  YK  + +
Sbjct:   491 SITEKLQRTAISDNTLSAESQLVDDEYF-QENPHVRPMVKGTAVAMQGPVYKGRNTMQPF 549

Query:   625 STASSENGAGK-PWKKHGNRNKKEKSRRL 652
               +++ + K P    G    + K+R+L
Sbjct:   550 QRRLNDDASPKYPMNAQGKPLSRRKARQL 578
```

Score = 47 (7.1 bits), Expect = 1.3e-60, Sum P(3) = 1.3e-60
Identities = 21/84 (25%), Positives = 35/84 (41%)

```
Query:   552 GRDPVTFQHQHQRLLENKMNSDEIKMQLGRNKKAKQIENIVDKTFFHQENVRALTKGVQA 611
             G D  T++       +   + +DE  +   R    K +E I +K  F         TK
Sbjct:   248 GEDLETYESTSSNEIPESLQADENDVHSSRIATLKVLEGIFEK--FASTLPDGKTKMTFG 305

Query:   612 VMGYKPGSGVVTASTASSENGAGK 635
             ++GY P  G   +ST ++  G+ K
Sbjct:   306 LVGY-PNVG--KSSTINALVGSKK 326
```

Score = 43 (6.5 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111
Identities = 7/13 (53%), Positives = 9/13 (69%)

```
Query:   638 KKHGNRNKKEKSR 650
             KKH  +NK+ K R
Sbjct:   596 KKHNKKNKRSKQR 608
```

Pedant information for DKFZphtes3_8e24, frame 3

Report for DKFZphtes3_8e24.3

[LENGTH]	658
[MW]	75226.58
[pI]	5.86
[HOMOL]	SWISSPROT:YAWG_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN CHROMOSOME I. 5e-56
[FUNCAT]	99 unclassified proteins [S. cerevisiae, YGL099w] 3e-55
[FUNCAT]	r general function prediction [M. jannaschii, MJ1464] 1e-16
[FUNCAT]	08.16 extracellular transport [S. cerevisiae, YER006w] 3e-09
[PIRKW]	P-loop 1e-27
[PIRKW]	GTP binding 1e-27
[SUPFAM]	conserved hypothetical protein MG442 7e-08

[PROSITE]	ATP_GTP_A	1
[PROSITE]	MYRISTYL	3
[PROSITE]	AMIDATION	2
[PROSITE]	CAMP_PHOSPHO_SITE	1
[PROSITE]	CK2_PHOSPHO_SITE	19
[PROSITE]	TYR_PHOSPHO_SITE	2
[PROSITE]	PKC_PHOSPHO_SITE	10
[PROSITE]	ASN_GLYCOSYLATION	2
[KW]	Alpha_Beta	
[KW]	LOW_COMPLEXITY	4.56 %

```
SEQ  MGRRRAPAGGSLGRALMRHQTQRSRSHRHTDSWLHTSELNDGYDWGRLNLQSVTEQSSLD
SEG  .................xxxxxxxxxxxxx..............................
PRD  ccccccccccchhhhhhhhhhhccccccccccccccccccccccccchhhhhhhccccch

SEQ  DFLATAELAGTEFVAEKLNIKFVPAEARTGLLSFEESQRIKKLHEENKQFLCIPRRPNWN
SEG  ............................................................
PRD  hhhhhhhhhhheeeecccceeeeeeecccccchhhhhhhhhhhhhhhhhhcccccccccc

SEQ  QNTTPEELKQAEKDNFLEWRRQLVRLEEEQKLILTPFERNLDFWRQLWRVIERSDIVVQI
SEG  ............................................................
PRD  cccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhcceeeee

SEQ  VDARNPLLFRCEDLECYVKEMDANKENVILINKADLLTAEQRSAWAMYFEKEDVKVIFWS
SEG  ............................................................
PRD  ecccccccccchhhhhhhhhhhccccceeeeecccchhhhhhhhhhhhhhhhccceeeeec

SEQ  ALAGAIPLNGDSEEEANRDDRQSNTTEFGHSSFDQAEISHSESEHLPARDSPSLSENPTT
SEG  ............................................................
PRD  ccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccccc

SEQ  DEDDSEYEDCPEEEEDDWQTCSEEDGPKEEDCSQDWKESSTADSEARSRKTPQKRQIHNF
SEG  xxxxxxxxxxxxxxxxx...........................................
PRD  ccccccccccccccccccccccccccccccccccccccchhhhhhhhccccccccccccc

SEQ  SHLVSKQELLELFKELHTGRKVKDGQLTVGLVGYPNVGKSSTINTIMGNKKVSVSATPGH
SEG  ............................................................
PRD  ccccchhhhhhhhhhhhhhhccccceeeeeeccccccccceeeecccccceeeeeeccccc

SEQ  TKHFQTLYVEPGLCLCDCPGLVMPSFVSTKAEMTCSGILPIDQMRDHVPPVSLVCQNIPR
SEG  ............................................................
PRD  cceeeeeeecccceeeccccccccccchhhhhhhhccccccccccccccccceeeeeccch

SEQ  HVLEATYGINIITPREDEDPHRPPTSEELLTAYGYMRGFMTAHGQPDQPRSARYILKDYV
SEG  ............................................................
PRD  hhhhhhhcccccccccccccccchhhhhhhhhhhhhhhccccccccccchhhhhhhhhcc

SEQ  SGKLLYCHPPPGRDPVTFQHQHQRLLENKMNSDEIKMQLGRNKKAKQIENIVDKTFFHQE
SEG  ............................................................
PRD  ccceeeecccccccccccchhhhhhhhhhccccchhhhhhhhhcchhhhhhhhhhhhccccch

SEQ  NVRALTKGVQAVMGYKPGSGVVTASTASSENGAGKPWKKHGNRNKKEKSRRLYKHLDM
SEG  ..........................................................
PRD  hhhhhhceeeeeecccccceeeccccccccccccccccccccccchhhhhhhhhhccc
```

Prosite for DKFZphtes3_8e24.3

PS00001	264->268	ASN_GLYCOSYLATION	PDOC00001
PS00001	359->363	ASN_GLYCOSYLATION	PDOC00001
PS00004	410->414	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDOC00005
PS00005	348->351	PKC_PHOSPHO_SITE	PDOC00005
PS00005	378->381	PKC_PHOSPHO_SITE	PDOC00005
PS00005	448->451	PKC_PHOSPHO_SITE	PDOC00005
PS00005	493->496	PKC_PHOSPHO_SITE	PDOC00005
PS00005	531->534	PKC_PHOSPHO_SITE	PDOC00005
PS00005	541->544	PKC_PHOSPHO_SITE	PDOC00005
PS00005	649->652	PKC_PHOSPHO_SITE	PDOC00005
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	93->97	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	155->159	CK2_PHOSPHO_SITE	PDOC00006
PS00006	252->256	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	279->283	CK2_PHOSPHO_SITE	PDOC00006

PS00006	281->285	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	299->303	CK2_PHOSPHO_SITE	PDOC00006
PS00006	305->309	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	322->326	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00006	365->369	CK2_PHOSPHO_SITE	PDOC00006
PS00006	449->453	CK2_PHOSPHO_SITE	PDOC00006
PS00006	493->497	CK2_PHOSPHO_SITE	PDOC00006
PS00006	505->509	CK2_PHOSPHO_SITE	PDOC00006
PS00007	480->488	TYR_PHOSPHO_SITE	PDOC00007
PS00007	190->198	TYR_PHOSPHO_SITE	PDOC00007
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	432->438	MYRISTYL	PDOC00008
PS00008	620->626	MYRISTYL	PDOC00008
PS00009	1->5	AMIDATION	PDOC00009
PS00009	378->382	AMIDATION	PDOC00009
PS00017	393->401	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphtes3_8e24.3)

DKFZphtes3_8g11
----

group: testes derived

DKFZphtes3_8g11 encodes a novel proline-rich 939 amino acid protein without similarity to known proteins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop).
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown, prolin ritch protein

1 EST hit (from testis library)

Sequenced by MediGenomix

Locus: unknown

Insert length: 3100 bp
Poly A stretch at pos. 3056, polyadenylation signal at pos. 3041

```
   1 AGAGTCTTCC CTCAGCATAT TTTACGATAG AGAAGATCTT GTTCCAATGG
  51 AAGAAAGTGA GGACTCACAG AGTGATTCCC AGACAAGGAT TTCTGAGTCC
 101 CAACACTCCC TCAAGCCAAA TTATCTTTCC CAGGCCAAGA CTGACTTCTC
 151 AGAACAGTTC CAGTTGCTAG AAGATCTGCA GCTAAAAATA GCAGCAAAAC
 201 TCTTAAGGAG TCAAATACCC CCCGATGTGC CTCCACCTCT AGCTTCAGGT
 251 CTAGTCCTAA AATACCCTAT CTGCCTACAG TGTGGCCGAT GTTCAGGACT
 301 TAATTGCCAT CATAAATTAC AGACCACTTC GGGGCCTTAT CTTCTTATCT
 351 ATCCACAGCT CCACCTTGTA CGCACTCCTG AAGGCCATGG TGAGGTTCGG
 401 TTGCATCTTG GCTTTAGGCT GAGAATTGGG AAAAGATCCC AAATCTCAAA
 451 GTATCGTGAA AGAGATAGAC CCGTCATACG GAGAAGCCCT ATATCACCAT
 501 CACAAAGGAA AGCTAAAATC TATACTCAAG CTTCCAAGAG TCCTACTTCC
 551 ACAATAGATT TGCAGTCTGG GCCTTCCCAG TCCCCTGCTC CTGTACAAGT
 601 CTACATCAGG CGAGGACAAC GCAGCAGGCC TGACTTAGTA GAAAAGACAA
 651 AAACTAGAGC ACCTGGGCAC TATGAATTCA CTCAAGTTCA CAACCTACCA
 701 GAGAGTGACT CTGAAAGCAC TCAGAATGAA AAACGGGCTA AAGTGAGAAC
 751 CAAAAAGACC TCTGATTCAA AATATCCAAT GAAGAGAATC ACCAAGCGAC
 801 TTAGAAAACA CAGAAAGTTC TACACAAACA GTAGAACCAC AATAGAGAGT
 851 CCTTCTAGGG AATTAGCAGC CCATTTAAGA AGGAAGAGGA TTGGAGCAAC
 901 TCAGACAAGT ACTGCCTCTT TAAAAAGACA ACCTAAGAAA CCTTCCCAAC
 951 CCAAGTTCAT GCAACTGCTT TTTCAGAGCC TAAAGCGGGC ATTCCAAACA
1001 GCACACAGAG TTATAGCTTC TGTTGGGCGG AAGCCTGTGG ACGGGACAAG
1051 GCCAGACAAT TTGTGGGCAA GCAAAAACTA TTATCCAAAA CAAAATGCCA
1101 GGGACTATTG CTTACCAAGC AGTATCAAAA GAGACAAGAG GTCAGCTGAC
1151 AAGCTAACGC CAGCAGGCTC AACCATTAAG CAGGAGGACA TATTGTGGGG
1201 AGGAACGGTC CAGTGCAGAT CAGCTCAACA GCCAAGAAGA GCTTACTCTT
1251 TCCAACCCAG ACCTCTTCGA CTGCCCAAGC CCACAGATTC CCAAAGTGGT
1301 ATTGCTTTCC AAACTGCCTC AGTGGGGCAG CCTCTGAGAA CTGTTCAAAA
1351 GGACAGTAGT AGCAGATCAA AGAAAAACTT CTATAGAAAT GAAACCTCCA
1401 GCCAGGAGTC TAAGAACTTG TCCACACCAG GAACCAGAGT TCAGGCCCGA
1451 GGAAGAATCC TACCTGGTTC CCCTGTGAAG AGAACCTGGC ACCGACATCT
1501 TAAAGACAAA CTCACACACA AGGAGCATAA CCACCCCAGC TTCTATAGGG
1551 AGAGAACCCC ACGCGGTCCT TCTGAGAGAA CCCGTCATAA CCCCTCTTGG
1601 AGAAACCATC GCAGTCCCTC TGAGAGAAGC CAACGCAGTT CCTTGGAGAG
1651 AAGACATCAC AGTCCCTCTC AGAGGAGCCA CTGCAGTCCC TCTAGGAAAA
1701 ACCATTCCAG TCCTTCTGAG AGAAGCTGGC GCAGTCCGTC TCAGAGAAAT
1751 CACTGCAGTC CCCCCGAGAG GAGCTGTCAC AGTCTCTCTG AAAGGGGCCT
1801 TCACAGTCCC TCTCAGAGGA GCCATCGCGG TCCCTCTCAG AGAAGACATC
1851 ACAGTCCCTC AGAGAGAAGC CATCGCAGTC CCTCAGAGAG AAGCCATCGC
1901 AGTCCCTCTG AGAGAAGACA TCGCAGTCCC TCCCAGAGGA GCCATCGCGG
1951 TCCCTCAGAG AGAAGCCATT GCAGTCCCTC TGAGAGAAGA CATCGCAGTC
2001 CCTCTCAGAG GAGCCATCGT GGTCCCTCTG AGAGAAGACA TCACAGTCCC
2051 TCTAAGAGAA GCCATCGCAG TCCCGCTCGG AGGAGCCATC GCAGTCCCTC
2101 AGAGAGAAGC CATCACAGTC CCTCTGAGAG AAGCCATCAC AGTCCCTCTG
2151 AGAGAAGACA TCACAGTCCC TCTGAGAGAA GCCATTGCAG TCCCTCTGAG
2201 AGAAGCCATT GCAGTCCCTC TGAGAGAAGA CATCGCAGTC CCTCTGAGAG
2251 AAGACATCAC AGTCCCTCAG AGAAAAGCCA TCACAGTCCC TCTGAGAGAA
2301 GCCATCACAG TCCCTCTGAG AGAAGACGTC ACAGTCCCTT GGAGAGGAGC
2351 CGTCACAGTC TCTTGGAGAG GAGCCATCGC AGTCCCTCTG AGAGGAGATC
2401 TCACAGGTCC TTTGAGAGGA GCCATCGTAG GATTTCTGAG AGAAGTCACA
2451 GTCCCTCAGA GAAGAGCCAC CTCAGTCCCT TGGAAAGAAG CCGTTGCAGT
2501 CCCTCTGAGA GGAGAGGACA CAGTTCCTCT GGGAAAACCT GTCACAGTCC
2551 CTCTGAGAGA AGCCATCGCA GTCCCTCCGG GATGAGGCAA GGGAGGACCT
2601 CTGAGAGGAG CCATCGCAGT TCCTGTGAGA GAACCCGTCA CAGTCCCTCT
```

```
2651 GAGATGAGGC CAGGGAGGCC CTCTGGGAGG AACCATTGCA GTCCCTCTGA
2701 GAGGAGCCGA CGCAGTCCCC TTAAGGAGGG ACTCAAGTAC AGTTTCCCTG
2751 GAGAGAGGCC CAGCCATAGT TTGTCTAGAG ATTTCAAGAA TCAAACAACT
2801 CTCCTCGGGA CCACACATAA AAATCCCAAA GCAGGGCAAG TGTGGAGGCC
2851 TGAAGCTACT CGATGAGGCG AGGTCCGCCC CTATTATTCA TTGTCCTAAG
2901 TCTTCATCGT GCTGCCCTTT CCAGGCTTCT TTCCTGCTCA GCCACTGCCT
2951 CCAATTCCTG CGCCCCCAGC GTGGAAAGGC TTCCATTTCT CTCTACCGGG
3001 GGGGAGGCGG GTGAGAATGG GTCTGTAATT TCTCTAAGAT GAATAAAGGG
3051 GCAGTTAATT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAGG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 47 bp to 2863 bp; peptide length: 939
Category: similarity to unknown protein
Classification: unclassified
Prosite motifs: ATP_GTP_A (824-832)

```
  1 MEESEDSQSD SQTRISESQH SLKPNYLSQA KTDFSEQFQL LEDLQLKIAA
 51 KLLRSQIPPD VPPPLASGLV LKYPICLQCG RCSGLNCHHK LQTTSGPYLL
101 IYPQLHLVRT PEGHGEVRLH LGFRLRIGKR SQISKYRERD RPVIRRSPIS
151 PSQRKAKIYT QASKSPTSTI DLQSGPSQSP APVQVYIRRG QRSRPDLVEK
201 TKTRAPGHYE FTQVHNLPES DSESTQNEKR AKVRTKKTSD SKYPMKRITK
251 RLRKHRKFYT NSRTTIESPS RELAAHLRRK RIGATQTSTA SLKRQPKKPS
301 QPKFMQLLFQ SLKRAFQTAH RVIASVGRKP VDGTRPDNLW ASKNYYPKQN
351 ARDYCLPSSI KRDKRSADKL TPAGSTIKQE DILWGGTVQC RSAQQPRRAY
401 SFQPRPLRLP KPTDSQSGIA FQTASVGQPL RTVQKDSSSR SKKNFYRNET
451 SSQESKNLST PGTRVQARGR ILPGSPVKRT WHRHLKDKLT HKEHNHPSFY
501 RERTPRGPSE RTRHNPSWRN HRSPSERSQR SSLERRHHSP SQRSHCSPSR
551 KNHSSPSERS WRSPSQRNHC SPPERSCHSL SERGLHSPSQ RSHRGPSQRR
601 HHSPSERSHR SPSERSHRSP SERRHRSPSQ RSHRGPSERS HCSPSERRHR
651 SPSQRSHRGP SERRHHSPSK RSHRSPARRS HRSPSERSHH SPSERSHHSP
701 SERRHHSPSE RSHCSPSERS HCSPSERRHR SPSERRHHSP SEKSHHSPSE
751 RSHHSPSERR RHSPLERSRH SLLERSHRSP SERRSHRSFE RSHRRISERS
801 HSPSEKSHLS PLERSRCSPS ERRGHSSSGK TCHSPSERSH RSPSGMRQGR
851 TSERSHRSSC ERTRHSPSEM RPGRPSGRNH CSPSERSRRS PLKEGLKYSF
901 PGERPSHSLS RDFKNQTTLL GTTHKNPKAG QVWRPEATR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8g11, frame 2

TREMBL:AF061185_1 gene: "car90"; product: "cyst germination specific acidic repeat protein precursor"; Phytophthora infestans cyst germination specific acidic repeat protein precursor (car90) gene, complete cds., N = 1, Score = 457, P = 2.3e-39

TREMBL:AC004561_38 gene: "F16P2.41"; product: "putative proline-rich protein"; Arabidopsis thaliana chromosome II BAC F16P2 genomic sequence, complete sequence., N = 1, Score = 340, P = 4.2e-27

TREMBL:AF062655_1 product: "plenty-of-prolines-101"; Mus musculus plenty-of-prolines-101 mRNA, complete cds., N = 1, Score = 313, P = 3.6e-24

PIR:PN0099 son3 protein - human (fragment), N = 1, Score = 292, P = 1.2e-22

>TREMBL:AF061185_1 gene: "car90"; product: "cyst germination specific acidic repeat protein precursor"; Phytophthora infestans cyst germination specific acidic repeat protein precursor (car90) gene, complete cds.

```
          Length = 1,489

  HSPs:

 Score = 457 (68.6 bits), Expect = 2.3e-39, P = 2.3e-39
 Identities = 91/444 (20%), Positives = 239/444 (53%)

Query:   475 SPVKRTWHRHLKDKLTHKEHNHPSFY-RERTPRGPSERTRHNPSWRNHRSPSERSQRSSL 533
             +P + T +  +++  T+        ++    E TP  P+E T + P+      +P+E +  +S
Sbjct:   584 APTEETMYAPIEET-TYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAST 642

Query:   534 ERRHHSPSQRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQRSH 593
             E    ++P++ +   +P+ +     P+E +  +P++     +P E + ++ +E    ++P++ +
Sbjct:   643 EETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 702

Query:   594 RGPSQRRHHSPSERSHRSPSERSHRSPSERRHRSPSQRSHRGPSERSHCSPSERRHRSPS 653
              P++     + P+E +  +P+E +  +P+E     +P + +  GP+E +  +P+E     +P+
Sbjct:   703 YAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYGPTEETTYAPTEATTYAPT 762

Query:   654 QRSHRGPSERRHHSPSKRSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHHSPSERSH 713
             + +     P+E     + P+  +  +P   +  +P+E + ++P+E + ++P+E    + P+E +
Sbjct:   763 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEETT 822

Query:   714 CSPSERSHCSPSERRHRSPSERRHHSPSEKSHHSPSERSHHSPSERRRHSPLERSRHSLL 773
             +P+E +    P+E     +P+E    ++P+E++ ++P+E++ ++P+E     ++P  E + +
Sbjct:   823 YAPTEETPYEPTEETTYTPTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPT 882

Query:   774 ERSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRGHSSSGKTC 832
             E +   +P++    ++    E + +    E +++P+E++  +P E +    P+E    ++ + +T
Sbjct:   883 EETTYAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETT 942

Query:   833 HSPSERSHRSPSGMRQGRTSERSHRSSCERTRHSPSEMRPGRPSGRNHCSPSERSRRSPL 892
             ++P+E +  +P+        +E +  +  E T + P+E      P+     +P+E + +P+
Sbjct:   943 YAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPI 1002

Query:   893 KEGLKYSFPGERPSHSLSRDFKNQTT 918
             +E   Y+ P E   +++ + +   + T
Sbjct:  1003 EE-TTYA-PTEETTYAPAEETPYEPT 1026

 Score = 445 (66.8 bits), Expect = 4.5e-38, P = 4.5e-38
 Identities = 83/394 (21%), Positives = 212/394 (53%)

Query:   502 ERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSW 561
             E TP  P+E T + P+      +P+E +  +  E    ++P++ +  +P+ +     P+E +
Sbjct:   763 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEETT 822

Query:   562 RSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPS 621
              +P++     P E + ++ +E    ++P++ +   P+++   ++P+E +  +P+E +     P+
Sbjct:   823 YAPTEETPYEPTEETTYTPTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPT 882

Query:   622 ERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSH 681
             E     +P++ +    P+E +  + +E    +P++ +     P+E    + P++ +  +P     +
Sbjct:   883 EETTYAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETT 942

Query:   682 RSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPS 741
              +P+E + ++P+E + ++P+E    ++P+E +    P+E +  +P+E     +P+E    ++P
Sbjct:   943 YAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPI 1002

Query:   742 EKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
             E++ ++P+E + ++P+E    + P E + ++  E +  +P+E  ++ S E + +     E +
Sbjct:  1003 EETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETT 1062

Query:   801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSSC 860
             ++P+E++    P E +  +P+E    ++ + +T ++P+E +  +P+         +E +
Sbjct:  1063 YAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPT 1122

Query:   861 ERTRHSPSEMRPGRPSGRNHCSPSERSRRSPLKE 894
             E T ++P+E      P+     +P E +   P +E
Sbjct:  1123 EETTYAPTEETTYAPTEETMYAPIEETTYGPTEE 1156

 Score = 439 (65.9 bits), Expect = 2.0e-37, P = 2.0e-37
 Identities = 86/421 (20%), Positives = 223/421 (52%)

Query:   475 SPVKRTWHRHLKDKLTHKEHNHPSFY-RERTPRGPSERTRHNPSWRNHRSPSERSQRSSL 533
             +P + T +    +K T+        ++    E TP  P+E T + P+      +P+E +  +S
Sbjct:   848 APTEETTYAPT-EKTTYAPTEETTYAPTEETPYEPTEETTYAPTKETTYAPTEETTYAST 906

Query:   534 ERRHHSPSQRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQRSH 593
             E    ++P++ +   +P+ +     P+E +  +P++     +P E + ++ +E    ++P++ +
Sbjct:   907 EETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 966
```

```
Query:   594 RGPSQRRHHSPSERSHRSPSERSHRSPSERRHRSPSQRSHRGPSERSHCSPSERRHRSPS 653
               P++    + P+E +   +P+E +   +P+E    +P + +    P+E +   +P+E      P+
Sbjct:   967 YAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPT 1026

Query:   654 QRSHRGPSERRHHSPSKRSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHHSPSERSH 713
             + +    P+E   ++P++ +   +      +   +P+E + ++P+E + + P+E   ++P+E +
Sbjct:  1027 EETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETT 1086

Query:   714 CSPSERSHCSPSERRHRSPSERRHHSPSEKSHHSPSERSHHSPSERRRHSPLERSRHSLL 773
              +P+E +   +P+E      +P+E     ++P+E++ + P+E + ++P+E     ++P E + ++ +
Sbjct:  1087 YAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPI 1146

Query:   774 ERSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRGHSSSGKTC 832
             E +    P+E   ++     E + +      E   ++P+E++      P     +   +P+E     ++ + +T
Sbjct:  1147 EETTYGPTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETT 1206

Query:   833 HSPSERSHRSPSGMRQGRTSERSHRSSCERTRHSPSEMRPGRPSGRNHCSPSERSRRSPL 892
             ++P+E +   +P+          +E +   +   E T + P+E        P+        +P+E +   +P
Sbjct:  1207 YAPTEETTYAPTEETPYEPTEETTYAPTEETTYEPTEETTYAPTEETTYAPTEETTYAPT 1266

Query:   893 KE 894
             +E
Sbjct:  1267 EE 1268
```

Score = 439 (65.9 bits), Expect = 2.0e-37, P = 2.0e-37
Identities = 91/434 (20%), Positives = 232/434 (53%)

```
Query:   475 SPVKRTWHRHLKDKLTHKEHNHPSFY-RERTPRGPSERTRHNPSWRNHRSPSERSQRSSL 533
             +P + T +     +K T+       ++      E TP    P+E T + P+          +P+E +   +S
Sbjct:   440 APTEETTYAPT-EKTTYAPTEETTYAPTEETPYEPTEETTYAPTKETTYAPTEETTYAST 498

Query:   534 ERRHHSPSQRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQRSH 593
             E     ++P++ +   +P+ +      P+E +   +P++      +P E + ++ +E     ++P++ +
Sbjct:   499 EETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 558

Query:   594 RGPSQRRHHSPSERSHRSPSERSHRSPSERRHRSPSQRSHRGPSERSHCSPSERRHRSPS 653
               P++    + P+E +   +P+E +   +P+E      +P + +      P+E +   +P+E        P+
Sbjct:   559 YAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPT 618

Query:   654 QRSHRGPSERRHHSPSKRSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHHSPSERSH 713
             + +    P+E   ++P++ +   +      +   +P+E + ++P+E + + P+E   ++P+E +
Sbjct:   619 EETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETT 678

Query:   714 CSPSERSHCSPSERRHRSPSERRHHSPSEKSHHSPSERSHHSPSERRRHSPLERSRHSLL 773
              +P+E +   +P+E      +P+E     ++P+E++ + P+E + ++P+E     ++P E + ++ +
Sbjct:   679 YAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPI 738

Query:   774 ERSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRGHSSSGKTC 832
             E +    P+E   ++     E + +      E   ++P+E++      P      +   +P+E     ++ + +T
Sbjct:   739 EETTYGPTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETT 798

Query:   833 HSPSERSHRSPSGMRQGRTSERSHRSSCERTRHSPSEMRPGRPSGRNHCSPSERSRRSPL 892
             ++P+E +   +P          T E   +   + E T ++P+E   P   P+          +P+E +   +P
Sbjct:   799 YAPTEETTYAP-------TEETPYEPT-EETTYAPTEETPYEPTEETTYTPTEETTYAPT 850

Query:   893 KEGLKYSFPGERPSHS 908
             +E    Y+  P  E+ +++
Sbjct:   851 EE-TTYA-PTEKTTYA 864
```

Score = 437 (65.6 bits), Expect = 3.3e-37, P = 3.3e-37
Identities = 85/417 (20%), Positives = 223/417 (53%)

```
Query:   502 ERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSW 561
             E TP   P+E T + P+         +P+E +   +   E+   ++P++ +   +P+ +      P+E +
Sbjct:   419 EETPYEPTEETTYTPTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTEETT 478

Query:   562 RSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPS 621
              +P++      +P E + ++ +E     ++P++ +     P++      + P+E +   +P+E +   +P+
Sbjct:   479 YAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPT 538

Query:   622 ERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSH 681
             E        +P++ +      P+E +   +P+E        P++ +      P+E     ++P++ +   +P     +
Sbjct:   539 EETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETT 598

Query:   682 RSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPS 741
              +P+E + ++P+E + + P+E     ++P+E +   +P+E +   + +E      +P+E     ++P+
Sbjct:   599 YAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPA 658

Query:   742 EKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
             E++ + P+E + ++P+E     ++P E + ++    E +   +P+E   ++     E + +      E +
Sbjct:   659 EETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETT 718
```

```
Query:    801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSSC 860
              ++P+E++  +P E +  +P E   +  + +T ++P+E +  +P+       +E +
Sbjct:   719 YAPTEETTYAPTEETMYAPIEETTYGPTEETTYAPTEATTYAPTEETPYAPTEETTYEPT 778

Query:    861 ERTRHSPSEMRPGRPSGRNHCSPSERSRRSPLKEGLKYSFPGERPSHSLSRDFKNQTT 918
                T ++P+E     P+      +P+E +  +P +E   Y  P E  +++ + +   + T
Sbjct:   779 GETTYAPTEETTYAPTEETTYAPTEETTYAPTEE-TPYE-PTEETTYAPTEETPYEPT 834

 Score = 428 (64.2 bits), Expect = 3.1e-36, P = 3.1e-36
 Identities = 89/440 (20%), Positives = 228/440 (51%)

Query:    473 PGSPVKRTWHRHLKDKLTHKEHNHPSFYR-ERTPRGPSERTRHNPSWRNHRSPSERSQRS 531
              P  P + T +   K+  T+       ++   E T    P+E T + P+       P+E +  +
Sbjct:   470 PYEPTEETTYAPTKET-TYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYA 528

Query:    532 SLERRHHSPSQRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQR 591
                E   ++P++ +  +  +P+ +   +P+E +  +P++      P E + ++ +E    ++P++
Sbjct:   529 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEE 588

Query:    592 SHRGPSQRRHHSPSERSHRSPSERSHRSPSERRHRSPSQRSHRGPSERSHCSPSERRHRS 651
              +    P +    ++P+E +  +P+E +    P+E    +P++ +    P+E +  + +E    +
Sbjct:   589 TMYAPIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYASTEETTYA 648

Query:    652 PSQRSHRGPSERRHHSPSKRSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHHSPSER 711
              P++ +     P+E    + P++ +   +P    +  +P+E + ++P+E + ++P+E    ++P+E
Sbjct:   649 PTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEE 708

Query:    712 SHCSPSERSHCSPSERRHRSPSERRHHSPSEKSHHSPSERSHHSPSERRRHSPLERSRHS 771
              +    P+E +  +P+E      +P+E    ++P E++ +  P+E +  ++P+E    ++P E + ++
Sbjct:   709 TPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYGPTEETTYAPTEATTYAPTEETPYA 768

Query:    772 LLERSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRGHSSSGK 830
                E +    P+    ++    E + +    E +++P+E++  +P E +    P+E    ++ + +
Sbjct:   769 PTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEETTYAPTEE 828

Query:    831 TCHSPSERSHRSPSGMRQGRTSERSHRSSCERTRHSPSEMRPGRPSGRNHCSPSERSRRS 890
              T + P+E +  +P+          +E +  +  E+T ++P+E       P+       P+E +  +
Sbjct:   829 TPYEPTEETTYTPTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTEETTYA 888

Query:    891 PLKEGLKYSFPGERPSHSLSRD 912
              P KE   Y+ P E  +++ + +
Sbjct:   889 PTKE-TTYA-PTEETTYASTEE 908

 Score = 427 (64.1 bits), Expect = 4.0e-36, P = 4.0e-36
 Identities = 81/394 (20%), Positives = 213/394 (54%)

Query:    502 ERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSW 561
              E T  GP+E T + P+       +P+E +  +  E     + P+ +   +P+ +   +P+E +
Sbjct:   739 EETTYGPTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETT 798

Query:    562 RSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPS 621
              +P++     +P E + +  +E    ++P++ +   P++     ++P+E +  +P+E +  +P+
Sbjct:   799 YAPTEETTYAPTEETPYEPTEETTYAPTEETPYEPTEETTYTPTEETTYAPTEETTYAPT 858

Query:    622 ERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSH 681
              E+   +P++ +    P+E +    P+E    +P++ +    P+E    ++ ++ +  +P   +
Sbjct:   859 EKTTYAPTEETTYAPTEETPYEPTEETTYAPTKETTYAPTEETTYASTEETTYAPTEETT 918

Query:    682 RSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPS 741
              +P+E + +  P+E + ++P+E    ++P+E +  +P+E +  +P+E    +P+E   +  P+
Sbjct:   919 YAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPT 978

Query:    742 EKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
              E++ ++P+E + ++P+E    ++P+E + ++  E +  +P+E   +  E + +   E +
Sbjct:   979 EETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETT 1038

Query:    801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSSC 860
              ++P+E++  +  E +  +P+E    ++ + +T + P+E +  +P+          +E + +  +
Sbjct:  1039 YAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPT 1098

Query:    861 ERTRHSPSEMRPGRPSGRNHCSPSERSRRSPLKE 894
              E T ++P+E     P+       P+E +  +P +E
Sbjct:  1099 EETTYAPTEETTYAPAEETPYEPTEETTYAPTEE 1132

 Score = 424 (63.6 bits), Expect = 8.5e-36, P = 8.5e-36
 Identities = 81/394 (20%), Positives = 210/394 (53%)

Query:    502 ERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSW 561
              E T    P+E T + P+       +P+E +  +  E     + P++ +   +P+ +   +P+E +
Sbjct:   939 EETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETM 998
```

```
Query:   562 RSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPS 621
             +P +      +P E + ++ +E    + P++ +     P++     ++P+E +   + +E +   +P+
Sbjct:   999 YAPIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYASTEETTYAPT 1058

Query:   622 ERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSH 681
             E       +P++ +     P+E +   +P+E       +P++ +     P+E     ++P++ +   +PA  +
Sbjct:  1059 EETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETP 1118

Query:   682 RSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPS 741
               P+E + ++P+E + ++P+E      ++P E +     P+E +   +P+E       +P+E     ++P+
Sbjct:  1119 YEPTEETTYAPTEETTYAPTEETMYAPIEETTYGPTEETTYAPTEATTYAPTEETPYAPT 1178

Query:   742 EKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
             E++ + P+   + ++P+E     ++P E + ++    E +   +P+E    +    E + +    E +
Sbjct:  1179 EETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEETTYAPTEETT 1238

Query:   801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSSC 860
             + P+E++   +P E +   +P+E     ++ + +T ++P + +     P+          +E +  +
Sbjct:  1239 YEPTEETTYAPTEETTYAPTEETTYAPTEETMYAPIDETTYGPTEETTYAPTEATTYAPT 1298

Query:   861 ERTRHSPSEMRPGRPSGRNHCSPSERSRRSPLKE 894
             E T ++P+E      P+G     +P+E +   +P +E
Sbjct:  1299 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEE 1332

 Score = 422 (63.3 bits), Expect = 1.4e-35, P = 1.4e-35
 Identities = 84/407 (20%), Positives = 216/407 (53%)

Query:   502 ERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSW 561
             E T    P+E T + P+            P+E +   +   E      + P++ +   +P+ +      +P+E +
Sbjct:   795 EETTYAPTEETTYAPTEETPYEPTEETTYAPTEETPYEPTEETTYTPTEETTYAPTEETT 854

Query:   562 RSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPS 621
              +P+++      +P E + ++ +E     + P++ +     P++     ++P+E +   + +E +   +P+
Sbjct:   855 YAPTEKTTYAPTEETTYAPTEETPYEPTEETTYAPTKETTYAPTEETTYASTEETTYAPT 914

Query:   622 ERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSH 681
             E       +P++ +     P+E +   +P+E       +P++ +     P+E     ++P++ +   +PA  +
Sbjct:   915 EETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETP 974

Query:   682 RSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPS 741
               P+E + ++P+E + ++P+E      ++P E +   +P+E +   +P+E         P+E      ++P+
Sbjct:   975 YEPTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTEETTYAPT 1034

Query:   742 EKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
             E++ ++P+E + ++ +E     ++P E + ++    E +     P+E   ++     E + +     E +
Sbjct:  1035 EETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETT 1094

Query:   801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSSC 860
             ++P+E++   +P E +   +P+E      +  + +T ++P+E +   +P+               E +
Sbjct:  1095 YAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYGPT 1154

Query:   861 ERTRHSPSEMRPGRPSGRNHCSPSERSRRSPLKEGLKYSFPGERPSHS 908
             E T ++P+E        P+         +P+E +     P   E     Y+ P E +++
Sbjct:  1155 EETTYAPTEATTYAPTEETPYAPTEETTYEPTGE-TTYA-PTEETTYA 1200

 Score = 421 (63.2 bits), Expect = 1.8e-35, P = 1.8e-35
 Identities = 86/418 (20%), Positives = 219/418 (52%)

Query:   491 HKEHNHPSFYRERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSR 550
             H  H         E T    P+E T + P+             +P+E +   +   E      + P++ +   +P+
Sbjct:   376 HYAHIEKPCDTEVTMYAPTEETTYAPTEETTYAPTEETPYEPTEETTYTPTE 435

Query:   551 KNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHR 610
              +      +P+E +   +P++++      +P E + ++ +E    + P++ +     P++      ++P+E +
Sbjct:   436 ETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTEETTYAPTKETTYAPTEETTY 495

Query:   611 SPSERSHRSPSERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSK 670
             + +E +   +P+E      +P++ +     P+E +   +P+E       +P++ +     P+E      ++P++
Sbjct:   496 ASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 555

Query:   671 RSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHR 730
              +   +PA   +     P+E + ++P+E + ++P+E      ++P E +   +P+E +   +P+E
Sbjct:   556 ETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPY 615

Query:   731 SPSERRHHSPSEKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFE 790
               P+E     ++P+E++ ++P+E + ++ +E     ++P E + ++    E +     P+E    ++     E
Sbjct:   616 EPTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTE 675

Query:   791 RS-HRRISERSHSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQG 849
               + +     E +++P+E++   +P E +   +P+E      +  + +T ++P+E +   +P+
```

```
Sbjct:   676 ETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMY 735

Query:   850 RTSERSHRSSCERTRHSPSEMRPGRPSGRNHCSPSERSRRSPLKEGLKYSFPGERPSHS 908
               E +      E T ++P+E      P+       +P+E +    P  E    Y+ P E  +++
Sbjct:   736 APIEETTYGPTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGE-TTYA-PTEETTYA 792

 Score = 420 (63.0 bits), Expect = 2.3e-35, P = 2.3e-35
 Identities = 82/393 (20%), Positives = 206/393 (52%)

Query:   502 ERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSW 561
             E TP  P+E T + P+      +P+E +  + +E   ++P++ +  +P+ +    P+E +
Sbjct:   971 EETPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTEETT 1030

Query:   562 RSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPS 621
               +P++      +P E + ++ +E    ++P++ +     P++     + P+E +   +P+E +  +P+
Sbjct:  1031 YAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPT 1090

Query:   622 ERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSH 681
             E      +P++ +     P+E +   +P+E       P++ +     P+E     ++P++ +  +P    +
Sbjct:  1091 EETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETT 1150

Query:   682 RSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPS 741
               P+E + ++P+E + ++P+E    ++P+E +    P+   +   +P+E     +P+E     ++P+
Sbjct:  1151 YGPTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPT 1210

Query:   742 EKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
             E++ ++P+E + + P+E    ++P E + +    E +  +P+E   ++     E + +    E
Sbjct:  1211 EETTYAPTEETPYEPTEETTYAPTEETTYEPTEETTYAPTEETTYAPTEETTYAPTEETM 1270

Query:   801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSSC 860
             ++P +++     P E +  +P+E    ++ + +T ++P+E +    P+G        +E +  +
Sbjct:  1271 YAPIDETTYGPTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPT 1330

Query:   861 ERTRHSPSEMRPGRP--------SGRNHCSPSE 885
             E T ++P E  P  P             S    C+  E
Sbjct:  1331 EETTYAPMEETPYEPAEESTSTVSTEKPCNTEE 1363

 Score = 419 (62.9 bits), Expect = 3.0e-35, P = 3.0e-35
 Identities = 83/411 (20%), Positives = 215/411 (52%)

Query:   502 ERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSW 561
             E T    P+E T + P++      +P+E +        E    ++P++ +  +P+ +     +P E +
Sbjct:   947 EETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETT 1006

Query:   562 RSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPS 621
               +P++      +P E + +    +E    ++P++ +     P++     ++ +E +   +P+E +  +P+
Sbjct:  1007 YAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPA 1066

Query:   622 ERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSH 681
             E      P++ +     P+E +   +P+E       +P++ +     P+E     ++P++ +     P    +
Sbjct:  1067 EETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETT 1126

Query:   682 RSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPS 741
               +P+E + ++P+E + ++P E      + P+E +   +P+E +   +P+E     +P+E     + P+
Sbjct:  1127 YAPTEETTYAPTEETMYAPIEETTYGPTEETTYAPTEATTYAPTEETPYAPTEETTYEPT 1186

Query:   742 EKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERS 800
              ++ ++P+E + ++P+E    ++P E + ++   E +   P+E  ++   E + +    E +
Sbjct:  1187 GETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEETTYAPTEETTYEPTEETT 1246

Query:   801 HSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSSC 860
             ++P+E++  +P E +  +P+E    ++      +T + P+E +   +P+          +E +  +
Sbjct:  1247 YAPTEETTYAPTEETTYAPTEETMYAPIDETTYGPTEETTYAPTEATTYAPTEETPYAPT 1306

Query:   861 ERTRHSPSEMRPGRPSGRNHCSPSERSRRSPLKEGLKYSFPGERPSHSLSRD 912
             E T + P+          P+       +P+E +   +P++E    Y  P E   + ++S +
Sbjct:  1307 EETTYEPTGETTYAPTEETTYAPTEETTYAPMEE-TPYE-PAEESTSTVSTE 1356

 Score = 415 (62.3 bits), Expect = 8.0e-35, P = 8.0e-35
 Identities = 84/423 (19%), Positives = 218/423 (51%)

Query:   473 PGSPVKRTWHRHLKDKLTHKEHNHPSFYR-ERTPRGPSERTRHNPSWRNHRSPSERSQRS 531
             P  P + T +     K+  T+         ++    E T      P+E T + P+        P+E +  +
Sbjct:   878 PYEPTEETTYAPTKET-TYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYA 936

Query:   532 SLERRHHSPSQRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQR 591
                  E      ++P++ +  +P+ +    +P+E +   +P++        P E + ++ +E     ++P++
Sbjct:   937 PTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEE 996

Query:   592 SHRGPSQRRHHSPSERSHRSPSERSHRSPSERRHRSPSQRSHRGPSERSHCSPSERRHRS 651
               +    P +    ++P+E +   +P+E +      P+E      +P++ +       P+E +   + +E       +
```

```
Sbjct:   997 TMYAPIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYASTEETTYA 1056

Query:   652 PSQRSHRGPSERRHHSPSKRSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHHSPSER 711
             P++ +   P+E   + P++ +  +P   +  +P+E + ++P+E + ++P+E    ++P+E
Sbjct:  1057 PTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEE 1116

Query:   712 SHCSPSERSHCSPSERRHRSPSERRHHSPSEKSHHSPSERSHHSPSERRRHSPLERSRHS 771
             +    P+E +  +P+E     +P+E    ++P E++ + P+E + ++P+E    ++P E + ++
Sbjct:  1117 TPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYGPTEETTYAPTEATTYAPTEETPYA 1176

Query:   772 LLERSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSERRGHSSSGK 830
               E +    P+    ++    E + +     E +++P+E++  +P E +    P+E     ++ + +
Sbjct:  1177 PTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEETTYAPTEE 1236

Query:   831 TCHSPSERSHRSPSGMRQGRTSERSHRSSCERTRHSPSEMRPGRPSGRNHCSPSERSRRS 890
             T + P+E +   +P++          +E +  +   E T ++P +        P+      +P+E +  +
Sbjct:  1237 TTYEPTEETTYAPTEETTYAPTEETTYAPTEETMYAPIDETTYGPTEETTYAPTEATTYA 1296

Query:   891 PLKE 894
             P +E
Sbjct:  1297 PTEE 1300

 Score = 403 (60.5 bits), Expect = 1.6e-33, P = 1.6e-33
 Identities = 84/394 (21%), Positives = 213/394 (54%)

Query:   501 RERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERS 560
             RE T    PSE T + P       +P+E++       +E     +    + ++ +P+++     ++P+ER
Sbjct:   319 REETTAAPSEDTTYAPREVTPYAPTEKPY--DVEETTYVTEESTY-APTKSETNAPTERM 375

Query:   561 WRSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSP 620
               +   ++    C     E + ++ +E     ++P++ +       P++      ++P+E +       P+E +   +P
Sbjct:   376 HYAHIEKP-CDT-EVTMYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEETTYTP 433

Query:   621 SERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRS 680
             +E      +P++ +     P+E++   +P+E      +P++ +      P+E     ++P+K +   +P     +
Sbjct:   434 TEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTEETTYAPTKETTYAPTEET 493

Query:   681 HRSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSP 740
               + +E + ++P+E + ++P+E    + P+E +  +P+E +  +P+E      +P+E    ++P
Sbjct:   494 TYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 553

Query:   741 SEKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISER 799
             +E++ ++P+E + +  P+E     ++P  E + ++   E +   +P E   ++     E + +     E
Sbjct:   554 TEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYAPTEETTYAPAEET 613

Query:   800 SHSPSEKSHLSPLERSRCSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSS 859
              + P+E++   +P E +   +P+E     ++S+ +T ++P+E +   +P+       +E +   +
Sbjct:   614 PYEPTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAP 673

Query:   860 CERTRHSPSEMRPGRPSGRNHCSPSERSRRSPLKE 894
              E T ++P+E       P+        +P+E +    +P +E
Sbjct:   674 TEETTYAPTEETTYAPTEETTYAPTEETTYAPAEE 708

 Score = 398 (59.7 bits), Expect = 5.5e-33, P = 5.5e-33
 Identities = 84/402 (20%), Positives = 209/402 (51%)

Query:   475 SPVKRTWHRHLKDKLTHKEHNHPSFY-RERTPRGPSERTRHNPSWRNHRSPSERSQRSSL 533
             +P + T +   +++  T+        ++   E TP  P+E T + P+        +P+E +   +S
Sbjct:   992 APTEETMYAPIEET-TYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAST 1050

Query:   534 ERRHHSPSQRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGLHSPSQRSH 593
             E     ++P++ +   +P+ +    P+E +   +P++      +P E + ++ +E     ++P++ +
Sbjct:  1051 EETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 1110

Query:   594 RGPSQRRHHSPSERSHRSPSERSHRSPSERRHRSPSQRSHRGPSERSHCSPSERRHRSPS 653
               P++     + P+E +   +P+E +   +P+E      +P + +    GP+E +   +P+E      +P+
Sbjct:  1111 YAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYGPTEETTYAPTEATTYAPT 1170

Query:   654 QRSHRGPSERRHHSPSKRSHRSPARRSHRSPSERSHHSPSERSHHSPSERRHHSPSERSH 713
             + +     P+E     + P+   +   +P     +   +P+E + ++P+E + ++P+E      + P+E +
Sbjct:  1171 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEETT 1230

Query:   714 CSPSERSHCSPSERRHRSPSERRHHSPSEKSHHSPSERSHHSPSERRRHSPLERSRHSLL 773
             +P+E +     P+E       +P+E      ++P+E++ ++P+E + ++P +     + P E + ++
Sbjct:  1231 YAPTEETTYEPTEETTYAPTEETTYAPTEETTYAPTEETMYAPIDETTYGPTEETTYAPT 1290

Query:   774 ERSHRSPSERRSHRSFERSHRRISERSHSPSEKSHLSPLERSRCSPSERRGHSSSGKTCH 833
             E +   +P+E     +      E        E ++ P+ ++   +P E +   +P+E       ++     +T +
Sbjct:  1291 EATTYAPTEETPYAPTE-------ETTYEPTGETTYAPTEETTYAPTEETTYAPMEETPY 1343

Query:   834 SPSERSHRSPSGMRQGRTSERSHRSSCERTRHSPSEMRPGRPS 876
```

```
             P+E S  + S  +    T E +   + E T   PS+     P+
Sbjct:  1344 EPAEESTSTVSTEKPCNTEEFTDEPTDEPT-DEPSDEPTDEPT 1385

 Score = 368 (55.2 bits), Expect = 9.5e-30, P = 9.5e-30
 Identities = 79/386 (20%), Positives = 211/386 (54%)

Query:   524 PSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSER 583
             PS+ ++  + E   + P + +  +PS     +P E +  +P+++ +     E + + ++E
Sbjct:   303 PSDETEAPT-EGTTYVPREETTAAPSEDTTYAPREVTPYAPTEKPY--DVEETTY-VTEE 358

Query:   584 GLHSPSQRSHRGPSQRRHHSPSER------SHRSPSERSHRSPSERRHRSPSQRSHRGPS 637
              ++P++      P++R H++ E+      +  +P+E + +P+E     +P++ +   P+
Sbjct:   359 STYAPTKSETNAPTERMHYAHIEKPCDTEVTMYAPTEETTYAPTEETTYAPTEETTYAPT 418

Query:   638 ERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSHRSPSERSHHSPSERSH 697
             E +   P+E   +P++ +   P+E   ++P++++ +P   +  +P+E + + P+E +
Sbjct:   419 EETPYEPTEETTYTPTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTEETT 478

Query:   698 HSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPSEKSHHSPSERSHHSPS 757
             ++P++   ++P+E +  + +E +  +P+E   +P+E   + P+E++ ++P+E + ++P+
Sbjct:   479 YAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPT 538

Query:   758 ERRRHSPLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSR 816
             E    ++P E + ++  E +  +P+E   +   E + +   E +++P+E++ +P+E +
Sbjct:   539 EETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETT 598

Query:   817 CSPSERRGHSSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSSCERTRHSPSEMRPGRPS 876
              +P+E     ++ + +T + P+E + +P+       +E + +S E T ++P+E     P+
Sbjct:   599 YAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPA 658

Query:   877 GRNHCSPSERSRRSPLKEGLKYSFPGERPSHS 908
                  P+E +  +P +E  Y+ P E  +++
Sbjct:   659 EETPYEPTEETTYAPTEE-TTYA-PTEETTYA 688

 Score = 337 (50.6 bits), Expect = 2.1e-26, P = 2.1e-26
 Identities = 66/328 (20%), Positives = 170/328 (51%)

Query:   502 ERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSW 561
             E T   P+E T + P+     +P+E +  +  E  ++P++ +  +P+ +  +P+E +
Sbjct:  1059 EETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETP 1118

Query:   562 RSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPS 621
                P++    +P E + ++ +E  +++P + +  GP++  ++P+E +  +P+E +  +P+
Sbjct:  1119 YEPTEETTYAPTEETTYAPTEETMYAPIEETTYGPTEETTYAPTEATTYAPTEETPYAPT 1178

Query:   622 ERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSH 681
             E    P+  +   P+E +  +P+E   +P++ +   P+E   + P++ +  +P +  +
Sbjct:  1179 EETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEETTYAPTEETT 1238

Query:   682 RSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPS 741
               P+E + ++P+E + ++P+E   ++P+E + +P + +  P+E  +P+E  ++P+
Sbjct:  1239 YEPTEETTYAPTEETTYAPTEETTYAPTEETMYAPIDETTYGPTEETTYAPTEATTYAPT 1298

Query:   742 EKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFERSHRRIS---- 797
             E++ ++P+E + + P+    ++P E + ++  E + +P E   +  E S   +S
Sbjct:  1299 EETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPMEETPYEPAEESTSTVSTEKP 1358

Query:   798 ----ERSHSPSEKSHLSPLERSRCSPSE 821
                 E +  P+++    P +    P++
Sbjct:  1359 CNTEEFTDEPTDEPTDEPSDEPTDEPTD 1386

 Score = 333 (50.0 bits), Expect = 5.7e-26, P = 5.7e-26
 Identities = 63/320 (19%), Positives = 166/320 (51%)

Query:   502 ERTPRGPSERTRHNPSWRNHRSPSERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSW 561
             E T   P+E T + P+     +P+E +  +  E  ++P++ +  P+ +  +P+E +
Sbjct:  1075 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETT 1134

Query:   562 RSPSQRNHCSPPERSCHSLSERGLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPS 621
             +P++    +P E + +  +E  ++P++ +  P++   ++P+E +  P+ + +  P+
Sbjct:  1135 YAPTEETMYAPIEETTYGPTEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPT 1194

Query:   622 ERRHRSPSQRSHRGPSERSHCSPSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSH 681
             E     +P++ +   P+E +  +P+E   P++ +   P+E   + P++ +  +P +  +
Sbjct:  1195 EETTYAPTEETTYAPTEETTYAPTEETPYEPTEETTYAPTEETTYEPTEETTYAPTEETT 1254

Query:   682 RSPSERSHHSPSERSHHSPSERRHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPS 741
             +P+E + ++P+E + ++P +  + P+E + +P+E + +P+E  +P+E  + P+
Sbjct:  1255 YAPTEETTYAPTEETMYAPIDETTYGPTEETTYAPTEATTYAPTEETPYAPTEETTYEPT 1314

Query:   742 EKSHHSPSERSHHSPSERRRHSPLERSRHSLLERSHRSPSERRSHRSFERSHRRISERSH 801
```

```
                 ++ ++P+E + ++P+E   ++P+E + +   E S  + S  +   + E +    E +
Sbjct: 1315 GETTYAPTEETTYAPTEETTYAPMEETPYEPAEESTSTVSTEKPCNTEEFTDEPTDEPTD 1374

Query:  802 SPSEKSHLSPLERSRCSPSE 821
             PS++    P +    P++
Sbjct: 1375 EPSDEPTDEPTDEPTDLPTD 1394

 Score = 303 (45.5 bits), Expect = 9.6e-23, P = 9.6e-23
 Identities = 70/322 (21%), Positives = 170/322 (52%)

Query:  584 GLHSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPSERRHRSPSQRSHRGPSERSHCS 643
             G + PS  +    P++   + P E +  +PSE +  +P E     +P+++ +      E ++ +
Sbjct:  299 GGYEPSDETE-APTEGTTYVPREETTAAPSEDTTYAPREVTPYAPTEKPY-DVEETTYVT 356

Query:  644 PSERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSHRSPSERSHHSPSERSHHSPSER 703
               E    +P++      P+ER H++  ++    +    + +P+E + ++P+E + ++P+E
Sbjct:  357 --EESTYAPTKSETNAPTERMHYAHIEKPCDTEV--TMYAPTEETTYAPTEETTYAPTEE 412

Query:  704 RHHSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPSEKSHHSPSERSHHSPSERRRHS 763
              ++P+E +    P+E +  +P+E     +P+E    ++P+EK+ ++P+E + ++P+E    +
Sbjct:  413 TTYAPTEETPYEPTEETTYTPTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYE 472

Query:  764 PLERSRHSLLERSHRSPSERRSHRSFERS-HRRISERSHSPSEKSHLSPLERSRCSPSER 822
             P E + ++  + +  +P+E  ++ S E + +    E +++P+E++   P E +  +P+E
Sbjct:  473 PTEETTYAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEE 532

Query:  823 RGHSSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSSCERTRHSPSEMRPGRPSGRNHCS 882
               ++  +  +T ++P+E +  +P+       +E +     E T ++P+E      P+        +
Sbjct:  533 TTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETMYA 592

Query:  883 PSERSRRSPLKEGLKYSFPGERP 905
             P E +  +P +E   Y+   E P
Sbjct:  593 PIEETTYAPTEE-TTYAPAEETP 614

 Score = 151 (22.7 bits), Expect = 2.0e-06, P = 2.0e-06
 Identities = 45/198 (22%), Positives = 103/198 (52%)

Query:  716 PSERSHCSPSERRHRSPSERRHHSPSEKSHHSPSERSHHSPSERRRHSPLERSRHSLLER 775
             PS+ +  +P+E      P E    +PSE + ++P E + ++P+E+     +E + + + E
Sbjct:  303 PSDETE-APTEGTTYVPREETTAAPSEDTTYAPREVTPYAPTEKPYD--VEETTY-VTEE 358

Query:  776 SHRSPSERRSHRSFERSHRRISERS-------HSPSEKSHLSPLERSRCSPSERRGHSSS 828
             S  +P++  ++  ER H   E+         ++P+E++ +P E + +P+E  ++ +
Sbjct:  359 STYAPTKSETNAPTERMHYAHIEKPCDTEVTMYAPTEETTYAPTEETTYAPTEETTYAPT 418

Query:  829 GKTCHSPSERSHRSPSGMRQGRTSERSHRSSCERTRHSPSEMRPGRPSGRNHCSPSERSR 888
              +T + P+E +  +P+        +E +  +   E+T ++P+E      P+        P+E +
Sbjct:  419 EETPYEPTEETTYTPTEETTYAPTEETTYAPTEKTTYAPTEETTYAPTEETPYEPTEETT 478

Query:  889 RSPLKEGLKYSFPGERPSHSLSRD 912
              +P KE   Y+ P E +++ + +
Sbjct:  479 YAPTKE-TTYA-PTEETTYASTEE 500
```

Pedant information for DKFZphtes3_8g11, frame 2

Report for DKFZphtes3_8g11.2

```
[LENGTH]     954
[MW]         110063.05
[pI]         11.40
[PROSITE]    ATP_GTP_A      1
[KW]         Irregular
[KW]         LOW_COMPLEXITY    27.67 %

SEQ    ESSLSIFYDREDLVPMEESEDSQSDSQTRISESQHSLKPNYLSQAKTDFSEQFQLLEDLQ
SEG    ................xxxxxxxxxxx.................................
PRD    ccceeeccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhh

SEQ    LKIAAKLLRSQIPPDVPPPLASGLVLKYPICLQCGRCSGLNCHHKLQTTSGPYLLIYPQL
SEG    ............................................................
PRD    hhhhhhhhhhcccccccccccceeeeecceeeccccccccccccccccccceeeehhhh

SEQ    HLVRTPEGHGEVRLHLGFRLRIGKRSQISKYRERDRPVIRRSPISPSQRKAKIYTQASKS
SEG    ............................................................
PRD    hcccccccccceeecccceeecccccccccccccccceeeeecccccchhhhhhhccccc

SEQ    PTSTIDLQSGPSQSPAPVQVYIRRGQRSRPDLVEKTKTRAPGHYEFTQVHNLPESDSEST
```

```
SEG   ............................................................
PRD   ccccccccccccccccceeeeeeecccccchhhhhhcccccceeeeeecccccccccch

SEQ   QNEKRAKVRTKKTSDSKYPMKRITKRLRKHRKFYTNSRTTIESPSRELAAHLRRKRIGAT
SEG   ............................................................
PRD   hhhhhhhhhhccccccccccchhhhhhhhhhhccccccccccccchhhhhhhhhhhhhcc

SEQ   QTSTASLKRQPKKPSQPKFMQLLFQSLKRAFQTAHRVIASVGRKPVDGTRPDNLWASKNY
SEG   ............................................................
PRD   ccchhhhhcccccccccchhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccc

SEQ   YPKQNARDYCLPSSIKRDKRSADKLTPAGSTIKQEDILWGGTVQCRSAQQPRRAYSFQPR
SEG   ............................................................
PRD   ccccccccccccccccccccccccccccccccccccceeccccccccccccccccccccc

SEQ   PLRLPKPTDSQSGIAFQTASVGQPLRTVQKDSSSRSKKNFYRNETSSQESKNLSTPGTRV
SEG   ............................................................
PRD   cccccccccccccceeeecccccccceeeeecccccccccccccccccccccccccccee

SEQ   QARGRILPGSPVKRTWHRHLKDKLTHKEHNHPSFYRERTPRGPSERTRHNPSWRNHRSPS
SEG   .......................................................xxxxx
PRD   eeeccccccccccccccccccccccccccccceeeecccccccccccccccccccccc

SEQ   ERSQRSSLERRHHSPSQRSHCSPSRKNHSSPSERSWRSPSQRNHCSPPERSCHSLSERGL
SEG   xxxxxxxxxxxxxx..............xxxxxxxxxxxx....................
PRD   chhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   HSPSQRSHRGPSQRRHHSPSERSHRSPSERSHRSPSERRHRSPSQRSHRGPSERSHCSPS
SEG   ..........xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   ERRHRSPSQRSHRGPSERRHHSPSKRSHRSPARRSHRSPSERSHHSPSERSHHSPSERRH
SEG   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   HSPSERSHCSPSERSHCSPSERRHRSPSERRHHSPSEKSHHSPSERSHHSPSERRRHSPL
SEG   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   ERSRHSLLERSHRSPSERRSHRSFERSHRRISERSHSPSEKSHLSPLERSRCSPSERRGH
SEG   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx....................
PRD   hhhhhhhhhhhccccccccchhhhhhhhhhhhcccccccccccccccccccccccccccc

SEQ   SSSGKTCHSPSERSHRSPSGMRQGRTSERSHRSSCERTRHSPSEMRPGRPSGRNHCSPSE
SEG   .......xxxxxxxxxxxx.........................................
PRD   cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ   RSRRSPLKEGLKYSFPGERPSHSLSRDFKNQTTLLGTTHKNPKAGQVWRPEATR
SEG   ......................................................
PRD   ccccccccccceeecccccccccccccccccccccccccccccccccccccccc
```

Prosite for DKFZphtes3_8g11.2

PS00017	839->847	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphtes3_8g11.2)

DKFZphtes3_8g5
----

group: testes derived

DKFZphtes3_8g5 encodes a novel 544 amino acid protein nearly identical to human KIAA087 protein.

The novel protein is a new splice variant of KIAA087.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

KIAA087, alternative spliced

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp
No poly A stretch found, no polyadenylation signal found

```
   1 CCGACATCGG CCGTGTCTCC AGCACCTGCC GGCGGCTGCG CGAGCTGTGC
  51 CAGAGCAGCG GGAAGGTGTG GAAGGAGCAG TTCCGGGTGA GGTGACCTTC
 101 CCTTATGAAA CACTACAGCC CCACCGACTA CGTCAATTGG TTGGAAGAGT
 151 ATAAAGTTCG GCAAAAAGCT GGGTTAGAAG CGCGGAAGAT TGTAGCCTCG
 201 TTCTCAAAGA GGTTCTTTTC AGAGCACGTT CCTTGTAATG GCTTCAGTGA
 251 CATTGAGAAC CTTGAAGGAC CAGAGATTTT TTTTGAGGAT GAACTGGTGT
 301 GTATCCTAAA TATGGAAGGA AGAAAAGCTT TGACCTGGAA ATACTACGCA
 351 AAAAAAATTC TTTACTACCT GCGGCAACAG AAGATCTTAA ATAATCTTAA
 401 GGCCTTTCTT CAGCAGCCAG ATGACTATGA GTCGTATCTT GAAGGTGCTG
 451 TATATATTGA CCAGTACTGC AATCCTCTCT CCGACATCAG CCTCAAAGAC
 501 ATCCAGGCCC AAATTGACAG CATCGTGGAG CTTGTTTGCA AAACCCTTCG
 551 GGGCATAAAC AGTCGCCACC CCAGCTTGGC CTTCAAGGCA GGTGAATCAT
 601 CCATGATAAT GGAAATAGAA CTCCAGAGCC AGGTGCTGGA TGCCATGAAC
 651 TATGTCCTTT ACGACCAACT GAAGTTCAAG GGGAATCGAA TGGATTACTA
 701 TAATGCCCTC AACTTATATA TGCATCAGGT TTTGATTCGC AGAACAGGAA
 751 TCCCAATCAG CATGTCTCTG CTCTATTTGA CAATTGCTCG GCAGTTGGGA
 801 GTCCCACTGG AGCCTGTCAA CTTCCCAAGT CACTTCTTAT TAAGGTGGTG
 851 CCAAGGCGCA GAAGGGGCGA CCCTGGACAT CTTTGACTAC ATCTACATAG
 901 ATGCTTTTGG GAAAGGCAAG CAGCTGACAG TGAAAGAATG CGAGTACTTG
 951 ATCGGCCAGC ACGTGACTGC AGCACTGTAT GGGGTGGTCA ATGTCAAGAA
1001 GGTGTTACAG AGAATGGTGG GAAACCTGTT AAGCCTGGGG AAGCGGGAAG
1051 GCATCGACCA GTCATACCAG CTCCTGAGAG ACTCGCTGGA TCTCTATCTG
1101 GCAATGTACC CGGACCAGGT GCAGCTTCTC CTCCTCCAAG CCAGGCTTTA
1151 CTTCCACCTG GGAATCTGGC CAGAGAAGTC TTTCTGTCTT GTTTTGAAGG
1201 TGCTTGACAT CCTCCAGCAC ATCCAAACCC TAGACCCGGG GCAGCACGGG
1251 GCGGTGGGCT ACCTGGTGCA GCACACTCTA GAGCACATTG AGCGCAAAAA
1301 GGAGGAGGTG GGCGTAGAGG TGAAGCTGCG CTCCGATGAG AAGCACAGAG
1351 ATGTCTGCTA CTCCATCGGG CTCATTATGA AGCATAAGAG GTATGGCTAT
1401 AACTGTGTGA TCTACGGCTG GGACCCCACC TGCATGATGG GACACGAGTG
1451 GATCCGGAAC ATGAACGTCC ACAGCCTGCC GCACGGCCAC CACCAGCCTT
1501 TCTATAACGT GCTGGTGGAG GACGGCTCCT GTCGATACGC AGCCCAAGAA
1551 AACTTGGAAT ATAACGTGGA GCCTCAAGAA ATCTCACACC CTGACGTGGG
1601 ACGCTATTTC TCAGAGTTTA CTGGCACTCA CTACATCCCA AACGCAGAGC
1651 TGGAGATCCG GTATCCAGAA GATCTGGAGT TTGTCTATGA AACGGTGCAG
1701 AATATTTACA GTGCAAAGAA AGAGAACATA GATGAGTAAA GTCTAGAGAG
1751 GACATTGCAC CTTTGCTGCT GCTGCTATCT TCCAAGAGAA CGGGACTCCG
1801 GAAGAAGACG TCTCCACGGA GCCCTCGGGA CCTGCTGCAC CAGGAAAGCC
1851 ACTCCACCAG TAGTGCTGGT TGCCTCCTAC TAAGTTTAAA TACCGTGTGC
1901 TCTTCCCCAG CTGCAAAGAC AATGTTGCTC TCCGCCTACA CTAGTGAATT
1951 AATCTGAAAG GCACTGTGTC AGTGGCATGG CTTGTATGCT TGTCCTGTGG
2001 TGACAGTTTG TGACATTCTG TCTTCATGAG GTCTCACAGT CGACGCTCCT
2051 GTAATCATTC TTTGTATTCA CTCCATTCCC CTGTCTGTCT GCATTTGTCT
2101 CAGAACATTT CCTTGGCTGG ACAGATGGGG TTATGCATTT GCAATAATTT
2151 CCTTCTGATT TCTCTGTGGA ACGTGTTCGG TCCCGAGTGA GGACTGTGTG
2201 TCTTTTTACC CTGAAGTTAG TTGCATATTC AGAGGTAAAG TTGTGTGCTA
2251 TCTTGGCAGC ATCTTAGAGA TGGAGACATT AACAAGCTAA TGGTAATTAG
2301 AATCATTTGA ATTTATTTTT TTCTAATATG TGAAACACAG ATTTCAAGTG
2351 TTTTATCTTT TTTTTTTTTA AATTTAAATG GGAATATAAC ACAGTTTTCC
2401 CTTCCATATT CCTCTCTTGA GTTTATGCAC ATCTCTATAA ATCATTAGTT
2451 TTCTATTTTA TTACATAAAA TTCTTTTAGA AAATGCAAAT AGTGAACTTT
2501 GTGAATGGAT TTTTCCATAC TCATCTACAA TTCCTCCATT TTAAATGACT
2551 ACTTTTATTT TTTAATTTAA AAAATCTACT TCAGTATCAT GAGTAGGTCT
2601 TACATCAGTG ATGGGTTCTT TTTGTAGTGA GACATACAAA TCTGATGTTA
```

```
2651 ATGTTTGCTC TTAGAAGTCA TACTCCATGG TCTTCAAAGA CCAAAAAATG
2701 AGGTTTTGCT TTTGTAATCA GGAAAAAAAA AATTAATGAA CCTTAAAAAA
2751 AAAAAAAAAA GG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 105 bp to 1736 bp; peptide length: 544
Category: known protein
Classification: unclassified

```
  1 MKHYSPTDYV NWLEEYKVRQ KAGLEARKIV ASFSKRFFSE HVPCNGFSDI
 51 ENLEGPEIFF EDELVCILNM EGRKALTWKY YAKKILYYLR QQKILNNLKA
101 FLQQPDDYES YLEGAVYIDQ YCNPLSDISL KDIQAQIDSI VELVCKTLRG
151 INSRHPSLAF KAGESSMIME IELQSQVLDA MNYVLYDQLK FKGNRMDYYN
201 ALNLYMHQVL IRRTGIPISM SLLYLTIARQ LGVPLEPVNF PSHFLLRWCQ
251 GAEGATLDIF DYIYIDAFGK GKQLTVKECE YLIGQHVTAA LYGVVNVKKV
301 LQRMVGNLLS LGKREGIDQS YQLLRDSLDL YLAMYPDQVQ LLLLQARLYF
351 HLGIWPEKSF CLVLKVLDIL QHIQTLDPGQ HGAVGYLVQH TLEHIERKKE
401 EVGVEVKLRS DEKHRDVCYS IGLIMKHKRY GYNCVIYGWD PTCMMGHEWI
451 RNMNVHSLPH GHHQPFYNVL VEDGSCRYAA QENLEYNVEP QEISHPDVGR
501 YFSEFTGTHY IPNAELEIRY PEDLEFVYET VQNIYSAKKE NIDE
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8g5, frame 3

TREMBLNEW:AB020682_1 gene: "KIAA0875"; product: "KIAA0875 protein"; Homo sapiens mRNA for KIAA0875 protein, partial cds., N = 1, Score = 2832, P = 5.5e-295

```
>TREMBLNEW:AB020682_1 gene: "KIAA0875"; product: "KIAA0875 protein";  Homo
     sapiens mRNA for KIAA0875 protein, partial cds.
          Length = 621

  HSPs:

 Score = 2832 (424.9 bits), Expect = 5.5e-295, P = 5.5e-295
 Identities = 537/544 (98%), Positives = 537/544 (98%)

Query:     1 MKHYSPTDYVNWLEEYKVRQKAGLEARKIVASFSKRFFSEHVPCNGFSDIENLEGPEIFF 60
             MKHYSPTDYVNWLEEYKVRQKAGLEARKIVASFSKRFFSEHVPCNGFSDIENLEGPEIFF
Sbjct:    85 MKHYSPTDYVNWLEEYKVRQKAGLEARKIVASFSKRFFSEHVPCNGFSDIENLEGPEIFF 144

Query:    61 EDELVCILNMEGRKALTWKYYAKKILYYLRQQKILNNLKAFLQQPDDYESYLEGAVYIDQ 120
             EDELVCILNMEGRKALTWKYYAKKILYYLRQQKILNNLKAFLQQPDDYESYLEGAVYIDQ
Sbjct:   145 EDELVCILNMEGRKALTWKYYAKKILYYLRQQKILNNLKAFLQQPDDYESYLEGAVYIDQ 204

Query:   121 YCNPLSDISLKDIQAQIDSIVELVCKTLRGINSRHPSLAFKAGESSMIMEIELQSQVLDA 180
             YCNPLSDISLKDIQAQIDSIVELVCKTLRGINSRHPSLAFKAGESSMIMEIELQSQVLDA
Sbjct:   205 YCNPLSDISLKDIQAQIDSIVELVCKTLRGINSRHPSLAFKAGESSMIMEIELQSQVLDA 264

Query:   181 MNYVLYDQLKFKGNRMDYYNALNLYMHQVLIRRTGIPISMSLLYLTIARQLGVPLEPVNF 240
             MNYVLYDQLKFKGNRMDYYNALNLYMHQVLIRRTGIPISMSLLYLTIARQLGVPLEPVNF
Sbjct:   265 MNYVLYDQLKFKGNRMDYYNALNLYMHQVLIRRTGIPISMSLLYLTIARQLGVPLEPVNF 324

Query:   241 PSHFLLRWCQGAEGATLDIFDYIYIDAFGKGKQLTVKECEYLIGQHVTAALYGVVNVKKV 300
             PSHFLLRWCQGAEGATLDIFDYIYIDAFGKGKQLTVKECEYLIGQHVTAALYGVVNVKKV
Sbjct:   325 PSHFLLRWCQGAEGATLDIFDYIYIDAFGKGKQLTVKECEYLIGQHVTAALYGVVNVKKV 384

Query:   301 LQRMVGNLLSLGKREGIDQSYQLLRDSLDLYLAMYPDQVQLLLLQARLYFHLGIWPEKSF 360
             LQRMVGNLLSLGKREGIDQSYQLLRDSLDLYLAMYPDQVQLLLLQARLYFHLGIWPEK
```

```
Sbjct:   385 LQRMVGNLLSLGKREGIDQSYQLLRDSLDLYLAMYPDQVQLLLLQARLYFHLGIWPEK-- 442

Query:   361 CLVLKVLDILQHIQTLDPGQHGAVGYLVQHTLEHIERKKEEVGVEVKLRSDEKHRDVCYS 420
                  VLDILQHIQTLDPGQHGAVGYLVQHTLEHIERKKEEVGVEVKLRSDEKHRDVCYS
Sbjct:   443 -----VLDILQHIQTLDPGQHGAVGYLVQHTLEHIERKKEEVGVEVKLRSDEKHRDVCYS 497

Query:   421 IGLIMKHKRYGYNCVIYGWDPTCMMGHEWIRNMNVHSLPHGHHQPFYNVLVEDGSCRYAA 480
             IGLIMKHKRYGYNCVIYGWDPTCMMGHEWIRNMNVHSLPHGHHQPFYNVLVEDGSCRYAA
Sbjct:   498 IGLIMKHKRYGYNCVIYGWDPTCMMGHEWIRNMNVHSLPHGHHQPFYNVLVEDGSCRYAA 557

Query:   481 QENLEYNVEPQEISHPDVGRYFSEFTGTHYIPNAELEIRYPEDLEFVYETVQNIYSAKKE 540
             QENLEYNVEPQEISHPDVGRYFSEFTGTHYIPNAELEIRYPEDLEFVYETVQNIYSAKKE
Sbjct:   558 QENLEYNVEPQEISHPDVGRYFSEFTGTHYIPNAELEIRYPEDLEFVYETVQNIYSAKKE 617

Query:   541 NIDE 544
             NIDE
Sbjct:   618 NIDE 621
```

Pedant information for DKFZphtes3_8g5, frame 3

Report for DKFZphtes3_8g5.3

```
[LENGTH]        544
[MW]            63307.22
[pI]            5.82
[HOMOL]         TREMBL:AB020682_1 gene: "KIAA0875"; product: "KIAA0875 protein"; Homo sapiens
mRNA for KIAA0875 protein, partial cds. 0.0
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY      1.84 %

SEQ   MKHYSPTDYVNWLEEYKVRQKAGLEARKIVASFSKRFFSEHVPCNGFSDIENLEGPEIFF
SEG   ............................................................
PRD   cccccccccchhhhhhhhhhhhchhhhhhhhhhhhhhhcccccccccccccccccceee

SEQ   EDELVCILNMEGRKALTWKYYAKKILYYLRQQKILNNLKAFLQQPDDYESYLEGAVYIDQ
SEG   ............................................................
PRD   eeeeeeeeeecccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccceeecceeeeee

SEQ   YCNPLSDISLKDIQAQIDSIVELVCKTLRGINSRHPSLAFKAGESSMIMEIELQSQVLDA
SEG   ............................................................
PRD   ccccccccchhhhhhhhhhhhhhhhhcccccccccceeeecccchhhhhhhhhhhhhhh

SEQ   MNYVLYDQLKFKGNRMDYYNALNLYMHQVLIRRTGIPISMSLLYLTIARQLGVPLEPVNF
SEG   ............................................................
PRD   hhhhhccccccccccchhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhcccccccccc

SEQ   PSHFLLRWCQGAEGATLDIFDYIYIDAFGKGKQLTVKECEYLIGQHVTAALYGVVNVKKV
SEG   ............................................................
PRD   cceeeeeeccccccceeeeeeeeeeccccceeeeeehhhhhhhhhhhhhhhhhhhhhh

SEQ   LQRMVGNLLSLGKREGIDQSYQLLRDSLDLYLAMYPDQVQLLLLQARLYFHLGIWPEKSF
SEG   ............................................................
PRD   hhhhhcchhhhhhhhhcccccccchhhhhhhhhhhcchhhhhhhhhhhhhhhccccceee

SEQ   CLVLKVLDILQHIQTLDPGQHGAVGYLVQHTLEHIERKKEEVGVEVKLRSDEKHRDVCYS
SEG   ...................................xxxxxxxxxx............
PRD   ehhhhhhhhhhhhhhcccccccccchhhhhhhhhhhhhhhhhhhhheeeeeccccccceeeecc

SEQ   IGLIMKHKRYGYNCVIYGWDPTCMMGHEWIRNMNVHSLPHGHHQPFYNVLVEDGSCRYAA
SEG   ............................................................
PRD   ccchhhhhhhceeeeeccccccccchhhhhhhhhhhccccccccccceeeeeeccccceeee

SEQ   QENLEYNVEPQEISHPDVGRYFSEFTGTHYIPNAELEIRYPEDLEFVYETVQNIYSAKKE
SEG   ............................................................
PRD   hhhhhhhhcccccccccceeeecccccccccchhhhhhcchhhhhhhhhhhhhhccccc

SEQ   NIDE
SEG   ....
PRD   cccc
```

(No Prosite data available for DKFZphtes3_8g5.3)

(No Pfam data available for DKFZphtes3_8g5.3)

DKFZphtes3_8m10
----

group: nucleic acid management

DKFZphtes3_8m10 encodes a novel 221 amino acid protein with strong similarity to polyadenylate-binding proteins.

The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA.

The new protein can find application in modulation of mRNA translation and processing/stability.

strong similarity to polyadenylate-binding protein

frame shift at Bp 707-710

Sequenced by MediGenomix

Locus: unknown

Insert length: 2107 bp
Poly A stretch at pos. 2052, polyadenylation signal at pos. 2033

```
   1 CGGAAAGGTC GCGGCTTGTG TGCCTGCGGG CAGCCGTGCC GAGAATGAAC
  51 CCCAGCACCC CCAGCTACCC AACGGCCTCG CTCTACGTGG GGGACCTCCA
 101 CCCCGACGTG ACTGAGGCGA TGCTCTACGA GAAGTTCAGC CCGGCAGGGC
 151 CCATCCTCTC CATCCGGATC TGCAGGGACT TGATCACCAG CGGCTCCTCC
 201 AACTACGCGT ATGTGAACTT CCAGCATACG AAGGACGCGG AGCATGCTCT
 251 GGACACCATG AATTTTGATG TTATAAAGGG CAAGCCAGTA CGCATCATGT
 301 GGTCTCAGCG TGATCCATCA CTTCGAAAAA GTGGAGTGGG CAACATATTC
 351 GTTAAAAATC TGGATAAGTC CATTAATAAT AAAGCACTGT ATGATACAGT
 401 TTCTGCTTTT GGTAACATCC TTTCGTGTAA CGTGGTTTGT GATGAAAATG
 451 GTTCCAAGGG TTATGGATTT GTACACTTTG AGACACACGA AGCAGCTGAA
 501 AGAGCTATTA AAAAAATGAA CGGAATGCTC CTAAATGGTC GCAAAGTATT
 551 TGTTGGACAA TTTAAGTCTC GTAAAGAACG AGAAGCTGAA CTTGGAGCTA
 601 GGGCAAAAGA GTTCCCCAAT GTTTACATCA AGAATTTTGG AGAAGACATG
 651 GATGATGAGC GCCTTAAGGA TCTCTTTGGC AAGTTCGGGC CCGCCTTAAG
 701 TGTGAATTAA TGACCGATGA AAGTGGAAAA TCCAAAGGAT TTGGATTTGT
 751 AAGCTTTGAA AGGCATGAAG ATGCACAGAA AGCTGTAGAT GAGATGAATG
 801 GAAAGGAGCT CAATGGAAAA CAAATTTACG TTGGTCGAGC TCAGAAAAAA
 851 GTGGAACGGC AGACGGAACT TAAGCGCACA TTTGAACAGA TGAAGCAAGA
 901 TAGGATCACC AGATACCAGG TTGTTAATCT TTATGTGAAA AATCTTGATG
 951 ATGGTATTGA TGATGAACGT CTCCGGAAAG CGTTTTCTCC ATTTGGTACA
1001 ATCACTAGTG CAAAGGTTAT GATGGAAGGT GGTCGCAGCA AAGGGTTTGG
1051 TTTTGTATGT TTCTCCTCCC CAGAAGAAGC CACTAAAGCA GTTACAGAAA
1101 TGAACGGTAG AATTGTGGCC ACAAAGCCAT TGTATGTAGC TTTAGCTCAG
1151 CGCAAAGAAG AGCGCCAGGC TTACCTCACT AACGAGTATA TGCAGAGAAT
1201 GGCAAGTGTA CGAGCTGTGC CCAACCAGCG AGCACCTCCT TCAGGTTACT
1251 TCATGACAGC TGTCCCACAG ACTCAGAACC ATGCTGCATA CTATCCTCCT
1301 AGCCAAATTG CTCGACTAAG ACCAAGTCCT CGCTGGACTG CTCAGGGTGC
1351 CAGACCTCAT CCATTCCAAA ATAAGCCCAG TGCTATCCGC CCAGGTGCTC
1401 CTAGAGTACC ATTTAGTACT ATGAGACCAG CTTCTTCACA GGTTCCACGA
1451 GTCATGTCAA CGCAGCGTGT TGCTAACACA TCAACACAGA CAGTGGGTCC
1501 ACGTCCTGCA GCTGCTGCTG CTGCTGCAGC TACCCCTGCT GTGCGCACGG
1551 TTCCACGGTA TAAATATGCT GCGGGAGTTC GCAATCCTCA GCAACATCGT
1601 AATGCACAGC CACAAGTTAC AATGCAACAG CTTGCTGTTC ATGTACAAGG
1651 TCAGGAAACT TTGACTGCCT CCAGGTTGGC ATCTGCCCCT CCTCAAAAGC
1701 AAAAGCAAAT GTTAGGTGAA CGGCTCTTTC CTCTTATTCA AGCCATGCAC
1751 CCTACTCTTG CTGGGAAAAT CACTGGCATG TTGTTGGAGA TTGATAATTC
1801 AGAACTTCTT TATATGCTCG AGTCTCCAGA GTCACTCCGT TCTAAGGTTG
1851 ATGAAGCTGT AGCTGTACTA CAAGCCCACC AAGCTAAAGA GGCTACCCAG
1901 AAAGCAGTTA ACAGTGCTAC CGGTGTTCCA ACTGTTTAAA ATTGATCAGA
1951 GACCACGAAA AGAAATTTGT GCTTCACCGA AGAAAAATAT CTAAACATCG
2001 AGAAACTATG GGAAAAAAAA TTGCAAAATC TAAAATAAAA AATGCAAAAT
2051 CTAAAATAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2101 AAAAAGG
```

BLAST Results

Entry HSPOLYAB from database EMBL:
Human mRNA for polyA binding protein
Score = 5420, P = 0.0e+00, identities = 1162/1243

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 707 bp to 1936 bp; peptide length: 410
Category: strong similarity to known protein
Classification: unset
Prosite motifs: RNP_1 (10-18)
RNP_1 (112-120)

```
  1 LMTDESGKSK GFGFVSFERH EDAQKAVDEM NGKELNGKQI YVGRAQKKVE
 51 RQTELKRTFE QMKQDRITRY QVVNLYVKNL DDGIDDERLR KAFSPFGTIT
101 SAKVMMEGGR SKGFGFVCFS SPEEATKAVT EMNGRIVATK PLYVALAQRK
151 EERQAYLTNE YMQRMASVRA VPNQRAPPSG YFMTAVPQTQ NHAAYYPPSQ
201 IARLRPSPRW TAQGARPHPF QNKPSAIRPG APRVPFSTMR PASSQVPRVM
251 STQRVANTST QTVGPRPAAA AAAAATPAVR TVPRYKYAAG VRNPQQHRNA
301 QPQVTMQQLA VHVQGQETLT ASRLASAPPQ KQKQMLGERL FPLIQAMHPT
351 LAGKITGMLL EIDNSELLYM LESPESLRSK VDEAVAVLQA HQAKEATQKA
401 VNSATGVPTV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8m10, frame 2

PIR:DNHUPA polyadenylate-binding protein - human, N = 1, Score = 1931, P = 1.7e-199

PIR:I48718 poly(A) binding protein - mouse, N = 1, Score = 1928, P = 3.6e-199

>PIR:DNHUPA polyadenylate-binding protein - human
Length = 633

HSPs:

Score = 1931 (289.7 bits), Expect = 1.7e-199, P = 1.7e-199
Identities = 384/415 (92%), Positives = 394/415 (94%)

```
Query:    1 LMTDESGKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFE 60
            +MTDESGKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKR FE
Sbjct:  219 VMTDESGKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRKFE 278

Query:   61 QMKQDRITRYQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVMMEGGRSKGFGFVCFS 120
            QMKQDRITRYQ VNLYVKNLDDGIDDERLRK FSPFGTITSAKVMMEGGRSKGFGFVCFS
Sbjct:  279 QMKQDRITRYQGVNLYVKNLDDGIDDERLRKEFSPFGTITSAKVMMEGGRSKGFGFVCFS 338

Query:  121 SPEEATKAVTEMNGRIVATKPLYVALAQRKEERQAYLTNEYMQRMASVRAVPN------Q 174
            SPEEATKAVTEMNGRIVATKPLYVALAQRKEERQA+LTN+YMQRMASVRAVPN      Q
Sbjct:  339 SPEEATKAVTEMNGRIVATKPLYVALAQRKEERQAHLTNQYMQRMASVRAVPNPVINPYQ 398

Query:  175 RAPPSGYFMTAVPQTQNHAAYYPPSQIARLRPSPRWTAQGARPHPFQNKPSAIRPGAPRV 234
             APPSGYFM A+PQTQN AAYYPPSQ+A+LRPSPRWTAQGARPHPFQN P AIRP APR
Sbjct:  399 PAPPSGYFMAAIPQTQNRAAYYPPSQVAQLRPSPRWTAQGARPHPFQNMPGAIRPAAPRP 458

Query:  235 PFSTMRPASSQVPRVMSTQRVANTSTQTVGPRPAAAAAAAATPAVRTVPRYKYAAGVRNP 294
            PFSTMRPASSQVPRVMSTQRVANTSTQT+GPRPAAAAAAA TPAVRTVP+YKYAAGVRNP
Sbjct:  459 PFSTMRPASSQVPRVMSTQRVANTSTQTMGPRPAAAAAAA-TPAVRTVPQYKYAAGVRNP 517

Query:  295 QQHRNAQPQVTMQQLAVHVQGQETLTASRLASAPPQKQKQMLGERLFPLIQAMHPTLAGK 354
            QQH NAQPQVTMQQ AVHVQGQE LTAS LASAPPQ+QKQMLGERLFPLIQAMHPTLAGK
Sbjct:  518 QQHLNAQPQVTMQQPAVHVQGQEPLTASMLASAPPQEQKQMLGERLFPLIQAMHPTLAGK 577

Query:  355 ITGMLLEIDNSELLYMLESPESLRSKVDEAVAVLQAHQAKEATQKAVNSATGVPTV 410
            ITGMLLEIDNSELL+MLESPESLRSKVDEAVAVLQAHQAKEA QKAVNSATGVPTV
Sbjct:  578 ITGMLLEIDNSELLHMLESPESLRSKVDEAVAVLQAHQAKEAAQKAVNSATGVPTV 633
```

Score = 315 (47.3 bits), Expect = 1.9e-27, P = 1.9e-27

```
 Identities = 71/163 (43%), Positives = 102/163 (62%)

Query:     1 LMTDESGKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFE 60
             ++ DE+G SKG+GFV FE  E A++A+++MNG  LN ++++VGR + + ER+ EL    +
Sbjct:   130 VVCDENG-SKGYGFVHFETQEAAERAIEKMNGMLLNDRKVFVGRFKSRKEREAELGARAK 188

Query:    61 QMKQDRITRYQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVMM-EGGRSKGFGFVCF 119
             +           N+Y+KN  + +DDERL+  F P     S KVM  E G+SKGFGFV F
Sbjct:   189 EF----------TNVYIKNFGEDMDDERLKDLFGP---ALSVKVMTDESGKSKGFGFVSF 235

Query:   120 SSPEEATKAVTEMNGRIVATKPLYVALAQRKEERQAYLTNEYMQ 163
                E+A KAV EMNG+ +  K +YV  AQ+K ERQ  L  ++ Q
Sbjct:   236 ERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRKFEQ 279

 Score = 214 (32.1 bits), Expect = 1.9e-14, P = 1.9e-14
 Identities = 50/150 (33%), Positives = 87/150 (58%)

Query:     8 KSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFEQMKQDRI 67
             +S G+ +V+F++  DA++A+D MN   + GK + +  +Q    R   L+++
Sbjct:    50 RSLGYAYVNFQQPADAERALDTMNFDVIKGKPVRIMWSQ----RDPSLRKS--------- 96

Query:    68 TRYQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVMMEGGRSKGFGFVCFSSPEEATK 127
                 V N+++KNLD  ID++ L   FS FG I S KV+ +   SKG+GFV F + E A +
Sbjct:    97 ---GVGNIFIKNLDKSIDNKALYDTFSAFGNILSCKVVCDENGSKGYGFVHFETQEAAER 153

Query:   128 AVTEMNGRIVATKPLYVALAQRKEERQAYL 157
             A+ +MNG ++  + ++V  + ++ER+A L
Sbjct:   154 AIEKMNGMLLNDRKVFVGRFKSRKEREAEL 183

 Score = 120 (18.0 bits), Expect = 4.8e-04, P = 4.8e-04
 Identities = 30/99 (30%), Positives = 54/99 (54%)

Query:    70 YQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVM--MEGGRSKGFGFVCFSSPEEATK 127
             Y + +LYV +L   + +  L + FSP G I S +V   M  RS G+ +V F  P +A +
Sbjct:     8 YPMASLYVGDLHPDVTEAMLYEKFSPAGPILSIRVCRDMITRRSLGYAYVNFQQPADAER 67

Query:   128 AVTEMNGRIVATKPLYVALAQRKEE-RQAYLTNEYMQRM 165
             A+  MN  ++  KP+ +  +QR   R++ + N +++ +
Sbjct:    68 ALDTMNFDVIKGKPVRIMWSQRDPSLRKSGVGNIFIKNL 106
```

Peptide information for frame 3

```
ORF from 45 bp to 707 bp; peptide length: 221
Category: strong similarity to known protein
Classification: unset
Prosite motifs: RNP_1 (138-146)
```

```
  1 MNPSTPSYPT ASLYVGDLHP DVTEAMLYEK FSPAGPILSI RICRDLITSG
 51 SSNYAYVNFQ HTKDAEHALD TMNFDVIKGK PVRIMWSQRD PSLRKSGVGN
101 IFVKNLDKSI NNKALYDTVS AFGNILSCNV VCDENGSKGY GFVHFETHEA
151 AERAIKKMNG MLLNGRKVFV GQFKSRKERE AELGARAKEF PNVYIKNFGE
201 DMDDERLKDL FGKFGPALSV N
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8m10, frame 3

SWISSPROT:PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING PROTEIN 1) (PABP 1)., N = 1, Score = 1039, P = 5.7e-105

PIR:I48718 poly(A) binding protein - mouse, N = 1, Score = 1031, P = 4e-104

PIR:DNHUPA polyadenylate-binding protein - human, N = 1, Score = 1009, P = 8.7e-102

```
>SWISSPROT:PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING
     PROTEIN 1) (PABP 1).
          Length = 636

  HSPs:
```

```
 Score = 1039 (155.9 bits), Expect = 5.7e-105, P = 5.7e-105
 Identities = 199/220 (90%), Positives = 205/220 (93%)

Query:     1 MNPSTPSYPTASLYVGDLHPDVTEAMLYEKFSPAGPILSIRICRDLITSGSSNYAYVNFQ 60
             MNPS PSYP ASLYVGDLHPDVTEAMLYEKFSPAGPILSIR+CRD+IT  S   YAYVNFQ
Sbjct:     1 MNPSAPSYPMASLYVGDLHPDVTEAMLYEKFSPAGPILSIRVCRDMITRRSLGYAYVNFQ 60

Query:    61 HTKDAEHALDTMNFDVIKGKPVRIMWSQRDPSLRKSGVGNIFVKNLDKSINNKALYDTVS 120
                DAE ALDTMNFDVIKGKPVRIMWSQRDPSLRKSGVGNIF+KNLDKSI+NKALYDT S
Sbjct:    61 QPADAERALDTMNFDVIKGKPVRIMWSQRDPSLRKSGVGNIFIKNLDKSIDNKALYDTFS 120

Query:   121 AFGNILSCNVVCDENGSKGYGFVHFETHEAAERAIKKMNGMLLNGRKVFVGQFKSRKERE 180
             AFGNILSC VVCDENGSKGYGFVHFET EAAERAI+KMNGMLLN RKVFVG+FKSRKERE
Sbjct:   121 AFGNILSCKVVCDENGSKGYGFVHFETQEAAERAIEKMNGMLLNDRKVFVGRFKSRKERE 180

Query:   181 AELGARAKEFPNVYIKNFGEDMDDERLKDLFGKFGPALSV 220
             AELGARAKEF NVYIKNFGEDMDDERLKDLFGKFGPALSV
Sbjct:   181 AELGARAKEFTNVYIKNFGEDMDDERLKDLFGKFGPALSV 220

 Score = 275 (41.3 bits), Expect = 4.1e-23, P = 4.1e-23
 Identities = 71/233 (30%), Positives = 120/233 (51%)

Query:     2 NPSTPSYPTASLYVGDLHPDVTEAMLYEKFSPAGPILSIRICRDLITSGSSNYAYVNFQH 61
             +PS        ++++ +L    +      LY+ FS  G ILS ++  D   S     + +   Q
Sbjct:    90 DPSLRKSGVGNIFIKNLDKSIDNKALYDTFSAFGNILSCKVVCDENGSKGYGFVHFETQE 149

Query:    62 TKD-AEHALDTMNFDVIKGKPVRIMW-SQRDPSL--RKSGVGNIFVKNLDKSINNKALYD 117
                + A   ++ M   +  K    R     +R+  L  R      N+++KN  + ++++ L D
Sbjct:   150 AAERAIEKMNGMLLNDRKVFVGRFKSRKEREAELGARAKEFTNVYIKNFGEDMDDERLKD 209

Query:   118 TVSAFGNILSCNVVCDENG-SKGYGFVHFETHEAAERAIKKMNGMLLNGRKVFVGQFKSR 176
                  FG   LS   V+ DE+G SKG+GFV FE HE A++A+ +MNG  LNG++++VG+ + +
Sbjct:   210 LFGKFGPALSVKVMTDESGKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKK 269

Query:   177 KEREAELGARAKEFP----------NVYIKNFGEDMDDERLKDLFGKFGPALS 219
               ER+ EL   + ++             N+Y+KN   + +DDERL+  F  FG   S
Sbjct:   270 VERQTELKRKFEQMKQDRITRYQGVNLYVKNLDDGIDDERLRKEFSPFGTITS 322

 Score = 227 (34.1 bits), Expect = 6.3e-18, P = 6.3e-18
 Identities = 57/187 (30%), Positives = 101/187 (54%)

Query:    12 SLYVGDLHPDVTEAMLYEKFSPAGPILSIRICRDLITSGSSNYAYVNFQHTKDAEHALDT 71
             ++Y+ +   D+ +  L + F   GP LS+++  D  +  S  + +V+F+  +DA+ A+D
Sbjct:   192 NVYIKNFGEDMDDERLKDLFGKFGPALSVKVMTDE-SGKSKGFGFVSFERHEDAQKAVDE 250

Query:    72 MNFDVIKGKPVRIMWSQR-----------------DPSLRKSGVGNIFVKNLDKSINNKA 114
             MN   + GK + +  +Q+                 D   R  GV N++VKNLD  I+++
Sbjct:   251 MNGKELNGKQIYVGRAQKKVERQTELKRKFEQMKQDRITRYQGV-NLYVKNLDDGIDDER 309

Query:   115 LYDTVSAFGNILSCNVVCDENGSKGYGFVHFETHEAAERAIKKMNGMLLNGRKVFVGQFK 174
             L     S FG I S   V+ +   SKG+GFV F + E A +A+ +MNG ++  + ++V   +
Sbjct:   310 LRKEFSPFGTITSAKVMMEGGRSKGFGFVCFSSPEEATKAVTEMNGRIVATKPLYVALAQ 369

Query:   175 SRKEREAEL 183
             ++ER+A L
Sbjct:   370 RKEERQAHL 378

 Score = 100 (15.0 bits), Expect = 2.3e-02, P = 2.3e-02
 Identities = 26/99 (26%), Positives = 53/99 (53%)

Query:     8 YPTASLYVGDLHPDVTEAMLYEKFSPAGPILSIRICRDLITSG-SSNYAYVNFQHTKDAE 66
             Y   +LYV +L   + +   L ++FSP G I S ++    ++   G S  + +V F   ++A
Sbjct:   291 YQGVNLYVKNLDDGIDDERLRKEFSPFGTITSAKV---MMEGGRSKGFGFVCFSSPEEAT 347

Query:    67 HALDTMNFDVIKGKPVRIMWSQRDPSLRKSGVGNIFVKNL 106
              A+  MN ++  KP+ + +  +QR  R++ + N +++ +
Sbjct:   348 KAVTEMNGRIVATKPLYVALAQRKEE-RQAHLTNQYMQRM 386
```

Pedant information for DKFZphtes3_8m10, frame 2

Report for DKFZphtes3_8m10.2

```
[LENGTH]     409
[MW]         45235.68
[pI]         10.08
[HOMOL]      SWISSPROT:PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING PROTEIN
1) (PABP 1). 0.0
```

[FUNCAT] 04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S. cerevisiae, YER165w] 1e-54
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER165w] 1e-54
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YER165w] 1e-54
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YER165w] 1e-54
[FUNCAT] 04.05.99 other mrna-transcription activities [S. cerevisiae, YNL016w] 1e-15
[FUNCAT] 11.01 stress response [S. cerevisiae, YGR159c] 1e-12
[FUNCAT] 04.01.04 rrna processing [S. cerevisiae, YGR159c] 1e-12
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YNL175c] 4e-09
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YPR112c] 5e-08
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YHR086w] 3e-07
[FUNCAT] 03.13 meiosis [S. cerevisiae, YHR086w] 3e-07
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YHR086w] 3e-07
[FUNCAT] 04.07 rna transport [S. cerevisiae, YOL123w HRP1 - CF Ib] 9e-07
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YCL011c] 3e-06
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YGR250c] 8e-06
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR432w] 2e-05
[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YDR432w] 2e-05
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YFR023w] 3e-05
[FUNCAT] 03.01 cell growth [S. cerevisiae, YBR212w] 3e-04
[BLOCKS] BL00030B Eukaryotic RNA-binding region RNP-1 proteins
[SCOP] d1sxl__ 4.34.7.1.3 Sex-lethal protein [(Drosophila melanogaster) 1e-17
[PIRKW] nucleus 0.0
[PIRKW] duplication 0.0
[PIRKW] RNA binding 0.0
[PIRKW] nucleolus 2e-09
[PIRKW] tandem repeat 2e-09
[PIRKW] single-stranded DNA binding 3e-06
[PIRKW] DNA binding 5e-13
[PIRKW] phosphoprotein 6e-10
[PIRKW] ribosome 3e-08
[PIRKW] mitochondrion 3e-08
[PIRKW] alternative splicing 9e-11
[PIRKW] chloroplast 2e-19
[PIRKW] transcription regulation 2e-07
[PIRKW] protein biosynthesis 3e-08
[SUPFAM] nucleolin 6e-10
[SUPFAM] glycine-rich RNA-binding protein 2e-07
[SUPFAM] unassigned ribonucleoprotein repeat-containing proteins 2e-19
[SUPFAM] polyadenylate-binding protein 0.0
[SUPFAM] ribonucleoprotein repeat homology 0.0
[PROSITE] RNP_1 2
[PFAM] RNA recognition motif. (aka RRM, RBD, or RNP domain)
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 5.62 %

```
SEQ    MTDESGKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQKKVERQTELKRTFEQ
SEG    ............................................................
1sxl-  ............................................................

SEQ    MKQDRITRYQVVNLYVKNLDDGIDDERLRKAFSPFGTITSAKVMMEGGRSKGFGFVCFSS
SEG    ............................................................
1sxl-  ...........CEEEECCCTTTTHHHHHHHHTTTTCCCCCEEECTTTCTTTEEEECTTT

SEQ    PEEATKAVTEMNGRIVATKPLYVALAQRKEERQAYLTNEYMQRMASVRAVPNQRAPPSGY
SEG    ............................................................
1sxl-  HHHHHHHHHHHHTTTCCCCCCCBCCBCC................................

SEQ    FMTAVPQTQNHAAYYPPSQIARLRPSPRWTAQGARPHPFQNKPSAIRPGAPRVPFSTMRP
SEG    ............................................................
1sxl-  ............................................................

SEQ    ASSQVPRVMSTQRVANTSTQTVGPRPAAAAAAAATPAVRTVPRYKYAAGVRNPQQHRNAQ
SEG    ...................xxxxxxxxxxxxxxxxxxxxxx...................
1sxl-  ............................................................

SEQ    PQVTMQQLAVHVQGQETLTASRLASAPPQKQKQMLGERLFPLIQAMHPTLAGKITGMLLE
SEG    ............................................................
1sxl-  ............................................................

SEQ    IDNSELLYMLESPESLRSKVDEAVAVLQAHQAKEATQKAVNSATGVPTV
SEG    .................................................
1sxl-  .................................................
```

Prosite for DKFZphtes3_8m10.2

```
PS00030        9->17   RNP_1               PDOC00030
PS00030      111->119  RNP_1               PDOC00030
```

Pfam for DKFZphtes3_8m10.2

```
HMM_NAME     RNA recognition motif. (aka RRM, RBD, or RNP domain)

HMM               *IYVGNLPWDtTEEDLrDlFsQFGpIvsIrMMrDReTGRSRGFAFVEFED
                   +YV+NL+   +++E LR +FS+FG I+S+++M+  E GRS+GF+FV F +
Query          74  LYVKNLDDGIDDERLRKAFSPFGTITSAKVMM--EGGRSKGFGFVCFSS   120

HMM               EEDAekAIdeMNGmeFmGRrIRV*
                  +E+A+KA+ EMNG+++   ++++V
Query         121 PEEATKAVTEMNGRIVATKPLYV    143
```

Pedant information for DKFZphtes3_8m10, frame 3

Report for DKFZphtes3_8m10.3

```
[LENGTH]       235
[MW]           26308.08
[pI]           8.95
[HOMOL]        SWISSPROT:PAB1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING PROTEIN
1) (PABP 1). 1e-113
[FUNCAT]       04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation)  [S.
cerevisiae, YER165w] 1e-64
[FUNCAT]       30.03 organization of cytoplasm      [S. cerevisiae, YER165w] 1e-64
[FUNCAT]       05.04 translation (initiation, elongation and termination)  [S. cerevisiae,
YER165w] 1e-64
[FUNCAT]       30.10 nuclear organization    [S. cerevisiae, YER165w] 1e-64
[FUNCAT]       03.19 recombination and dna repair    [S. cerevisiae, YFR023w] 1e-24
[FUNCAT]       11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair)        [S. cerevisiae, YFR023w] 1e-24
[FUNCAT]       04.05.99 other mrna-transcription activities        [S. cerevisiae, YNL016w]
2e-19
[FUNCAT]       04.05.03 mrna processing (splicing)  [S. cerevisiae, YOR319w] 2e-14
[FUNCAT]       04.01.04 rrna processing       [S. cerevisiae, YGR159c] 1e-11
[FUNCAT]       11.01 stress response  [S. cerevisiae, YGR159c] 1e-11
[FUNCAT]       99 unclassified proteins       [S. cerevisiae, YGR250c] 1e-09
[FUNCAT]       04.07 rna transport    [S. cerevisiae, YOL123w HRP1 - CF Ib] 1e-09
[FUNCAT]       30.13 organization of chromosome structure   [S. cerevisiae, YCL011c] 8e-09
[FUNCAT]       98 classification not yet clear-cut  [S. cerevisiae, YPR112c] 2e-08
[FUNCAT]       03.13 meiosis  [S. cerevisiae, YHR086w] 2e-08
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YBR212w] 3e-08
[FUNCAT]       03.01 cell growth      [S. cerevisiae, YBR212w] 3e-08
[FUNCAT]       06.04 protein targeting, sorting and translocation   [S. cerevisiae, YDR432w]
3e-04
[FUNCAT]       08.01 nuclear transport        [S. cerevisiae, YDR432w] 3e-04
[BLOCKS]       BL00030B Eukaryotic  RNA-binding region RNP-1 proteins
[BLOCKS]       BL00900D Bacteriophage-type RNA polymerase family proteins signatur
[SCOP]         d1sxl__ 4.34.7.1.3 Sex-lethal protein  [(Drosophila melanogaster)  9e-23
[SCOP]         d2u1a__ 4.34.7.1.2 U1A protein  [human (Homo sapiens)  6e-24
[SCOP]         d1up1_2 4.34.7.1.1 Nuclear ribonucleoprotein A1, RNP A1, UP 1e-13
[PIRKW]        nucleus 1e-110
[PIRKW]        duplication 1e-110
[PIRKW]        RNA binding 1e-110
[PIRKW]        nucleolus 4e-10
[PIRKW]        tandem repeat 4e-10
[PIRKW]        single-stranded DNA binding 1e-06
[PIRKW]        DNA binding 9e-12
[PIRKW]        phosphoprotein 4e-10
[PIRKW]        mitochondrion 6e-07
[PIRKW]        heterotrimer 4e-06
[PIRKW]        alternative splicing 1e-15
[PIRKW]        chloroplast 5e-11
[PIRKW]        transcription regulation 3e-09
[PIRKW]        GTP binding 2e-06
[SUPFAM]       helix-destabilizing protein 1e-07
[SUPFAM]       nucleolin 4e-10
[SUPFAM]       glycine-rich RNA-binding protein 2e-07
[SUPFAM]       yeast HRP1 protein 2e-08
```

```
[SUPFAM]       unassigned ribonucleoprotein repeat-containing proteins 3e-25
[SUPFAM]       polyadenylate-binding protein 1e-112
[SUPFAM]       ribonucleoprotein repeat homology 1e-112
[PROSITE]      RNP_1  1
[PFAM]         RNA recognition motif. (aka RRM, RBD, or RNP domain)
[KW]           All_Beta
[KW]           3D

SEQ     ERSRLVCLRAAVPRMNPSTPSYPTASLYVGDLHPDVTEAMLYEKFSPAGPILSIRICRDL
1hal-   ..........................EEEETTTTTTCHHHHHHHHHGGGCCEEEEEEEEETT

SEQ     ITSGSSNYAYVNFQHTKDAEHALDTMNFDVIKGKPVRIMWSQRDPSLRKSGVGNIFVKNL
1hal-   TTTCEEEEEEEEECCHHHHHHHHHHHTTEEE-TT---EEEEEEECTTTTCCCCCEEEEECC

SEQ     DKSINNKALYDTVSAFGNILSCNVVCDENGSKGYGFVHFETHEAAERAIKKMNGMLLNGR
1hal-   TTTTCHHHHHHHHHGGGCCEEEEEEEEETTTTTCEEEEEEECCHHHHHHHH...........

SEQ     KVFVGQFKSRKEREAELGARAKEFPNVYIKNFGEDMDDERLKDLFGKFGPALSVN
1hal-   ......................................................
```

Prosite for DKFZphtes3_8m10.3

```
PS00030     152->160    RNP_1                PDOC00030
```

Pfam for DKFZphtes3_8m10.3

```
HMM_NAME       RNA recognition motif. (aka RRM, RBD, or RNP domain)

HMM                *IYVGNLPWDtTEEDLrDlFsQFGpIvsIrMMrDReTGRSRGFAFVEFED
                    +YVG+L +D+TE +L + FS+ GPI+SIR+ RD  T  S  +A+V+F+
Query           27  LYVGDLHPDVTEAMLYEKFSPAGPILSIRICRDLITSGSSNYAYVNFQH    75

HMM                EEDAekAIdeMNGmeFmGRrIRV*
                    DAE A+D+MN  ++ G+++R+
Query           76 TKDAEHALDTMNFDVIKGKPVRI     98

HMM                *IYVGNLPWDtTEEDLrDlFsQFGpIvsIrMMrDReTGRSRGFAFVEFED
                    I+V+NL+ +++    L D  S FG I+S++++ D    + S+G++FV FE+
Query          115  IFVKNLDKSINNKALYDTVSAFGNILSCNVVCD--ENGSKGYGFVHFET    161

HMM                EEDAekAIdeMNGmeFmGRrIRV*
                   +E+AE+AI +MNGM+++GR++ V
Query          162 HEAAERAIKKMNGMLLNGRKVFV     184
```

DKFZphtes3_8p7

group: testes derived

DKFZphtes3_8p7 encodes a novel 412 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

2 EST hits (both from testis librarys)

Sequenced by MediGenomix

Locus: unknown

Insert length: 2899 bp
Poly A stretch at pos. 2870, polyadenylation signal at pos. 2852

```
   1 CCGACCCGCC CTGGGGTGCT GCGTGCGCTG CCTGCTCCCG CCTGAGGAAA
  51 ACACTGCCCA TGGCGCAAGG CCGGGAGCGC GACGAAGGCC CCCACTCCGC
 101 CGGCGGCGCG TCCTTGTCCG TGAGATGGGT GCAAGGATTC CCTAAGCAGA
 151 ATGTTCATTT GTCAACGACA ACACCATTTG CTACCCTTGT GGGAATTATG
 201 TAATATTTAT TAATATTGAA ACCAAGAAAA AGACTGTACT GCAGTGTAGT
 251 AATGGAATTG TGGGCGTCAT GGCAACTAAC ATCCCCTGTG AAGTTGTGGC
 301 TTTTTCTGAC CGGAAGCTAA AACCTCTCAT CTACGTATAC AGCTTTCCAG
 351 GATTGACCAG AAGGACCAAA TTGAAAGGCA ACATTCTCCT GGACTACACT
 401 TTACTTTCAT TCAGTTACTG TGGCACCTAC CTGGCTAGTT ACTCCTCTCT
 451 CCCAGAATTT GAACTGGCCC TTTGGAACTG GGAATCGAGT ATCATTTTGT
 501 GTAAGAAATC ACAGCCTGGA ATGGATGTGA ACCAAATGTC TTTTAACCCC
 551 ATGAACTGGC GCCAGCTGTG CTTATCAAGT CCAAGTACAG TGAGCGTGTG
 601 GACCATTGAA AGAAGTAACC AGGAGCATTG TTTCAGAGCA AGGTCGGTGA
 651 AATTACCTCT AGAAGATGGG TCATTTTTTA ATGAAACGGA TGTCGTTTTC
 701 CCCCAGTCGT TGCCGAAAGA TCTCATCTAT GGTCCCGTGC TGCCACTGTC
 751 AGCCATTGCC GGGCTGGTAG GCAAAGAGGC AGAGACTTTC CGGCCGAAAG
 801 ATGATCTATA TCCTTTGCTT CACCCGACTA TGCATTGCTG GACTCCAACA
 851 AGTGACTTGT ACATTGGCTG TGAAGAGGGT CATCTTTTAA TGATTAATGG
 901 AGACACCTTG CAAGTGACTG TACTTAATAA GATAGAAGAG GAATCGCCAT
 951 TGGAAGACAG AAGAAATTTT ATCAGTCCAG TAACCTTGGT ATATCAGAAG
1001 GAGGGCGTGC TGGCTTCTGG AATTGATGGC TTTGTGTATT CTTTTATTAT
1051 TAAAGATAGA AGTTACATGA TCGAGGATTT TCTTGAGATT GAAAGACCTG
1101 TAGAACATAT GACATTTTCT CCCAATTATA CAGTGTTGCT GATTCAAACA
1151 GACAAGGGAT CTGTTTATAT CTACACTTTT GGTAAGGAGC CAACCTTAAA
1201 TAAAGTCCTA GATGCTTGTG ATGGGAAATT TCAGGCAATT GACTTTATCA
1251 CACCTGGAAC CCAATACTTC ATGACACTTA CATATTCAGG GGAAATTTGT
1301 GTTTGGTGGC TGGAGGATTG TGCTTGTGTA AGCAAGATTT ATCTGAATAC
1351 CCTAGCAACG GTTCTGGCTT GCTGTCCATC CTCCCTCTCT GCAGCCGTGG
1401 GCACGGAGGA TGGCTCGGTC TACTTCATCA GCGTATATGA TAAGGAATCC
1451 CCTCAGGTCG TGCACAAGGC CTTTCTCTCG GAATCGTCCG TGCAGCACGT
1501 CGTGTAAGTC CTTTCTGCCT CCAGGAGCGG CTCCGTGTCA CACCCGTCTG
1551 TTGAAAATTC TAGTGAAGCC ATCCTTTCTT TTAATTTTAA GTTTTACGTG
1601 TTTCATTTGT TTTGAATGTT AATATATTCA CACAGTTCAA CACTCAAAAG
1651 GTACAGAGGG CTGTGTAGTA AAGTACCCCC CATACCCAGG TCTGTCCTTG
1701 CAGGCAGCCT GGTACCAATT TCTCATGTCT CTCCTGAGAT GTTTTATCCA
1751 TGAACAAGCA AAACATAATA AGCACTTCTT TTTACTTGTA TCAATGGCCA
1801 TCATGTGTGT ATAGTGTGCC AGGCACTTCT GCTGTATTAA CTCCATGAGG
1851 TAAACACTCT TGTTGTCTCT ATTTGACAGG TGAGGAAGAT AAGGCACAAG
1901 GATTTTAAAT AACTTGCTCA ATAGTACACA GATAGTGAAT GGCAAATGTT
1951 GGGATTTGAA CCCAGGTAGT TGGGCTGCAG AGTCACTGCC TTTGCTCTTA
2001 AAAGGAGAAA ACTATGTACA ATGCCTCATT TCTTTTTTCA CTTAATCGTA
2051 TATCTTGGAG AATGTTTTAT ATCCACACAT AAAGACCAGC CTGATTATTT
2101 GTATAGCCAC ATAGTATTCC ATTATATGAA TATACTATCA TTTTTTAAAA
2151 ACGGTATATT AATGAACATT TAGAGTATTT CAAAACTTTT GAAGCAATAC
2201 TTTTAAGATG ATAATATAGA GACATTAGAT TTGGACTTGT AGGTGCTATC
2251 ATTATTACTG TTTCTTTTTA ATTTATTATA TTATTAGGTA TTAATAAGAA
2301 CAGACATTTG TATTCTGCTT TACAGCTTGA GATCACTGTA GCTTGTGGCA
2351 TGTGATCCTC AAAACACCAG TCAGAAAGGT GTTATTCTTA TCCCTATTAG
2401 ACAAATTAGG GAATTCAGGG TTAGAGAGGT GAGGAAAAGC ATTGTCCAAG
2451 ATTACACATT ACACAGCTAG CACACTGAGG AGCTGGCCCT GCCACTGTGG
2501 ACTGCCCAGC TCCACCACCC TAGCTCAGTG GGGAAGGATG GATAACCTCC
2551 TTCCATTTAC CCCCTGCCTT TCTGCACTGT CATTTTTTTG TGCCTTTCCT
2601 TTCTCAGATC CTCTTATTCT AATTTACATC TTCCCACTTT TTCTAATTTG
2651 ATAAAGTTGT AGACATGTTT CACTACATTC TTCCTCCCAC TGCCAGGTAC
2701 CAGACACAGG GTAATGAAAT GTCACACCCA CCACTAATTT GAGAATTGCT
```

```
2751 TATTTGCGCT TGAAACATCA AGAAAGCTCT ACCGACAGAC ATGTTTCATT
2801 CACTTATGAT GAACCAACTG CCCATCTTTA CTGAATCTTC TTGACTGTAT
2851 TTATTAAAGT TGCAATTTGG AAATAAAAAA AAAAAAAAAA AAAAAAAGG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 269 bp to 1504 bp; peptide length: 412
Category: putative protein
Classification: no clue

```
  1 MATNIPCEVV AFSDRKLKPL IYVYSFPGLT RRTKLKGNIL LDYTLLSFSY
 51 CGTYLASYSS LPEFELALWN WESSIILCKK SQPGMDVNQM SFNPMNWRQL
101 CLSSPSTVSV WTIERSNQEH CFRARSVKLP LEDGSFFNET DVVFPQSLPK
151 DLIYGPVLPL SAIAGLVGKE AETFRPKDDL YPLLHPTMHC WTPTSDLYIG
201 CEEGHLLMIN GDTLQVTVLN KIEEESPLED RRNFISPVTL VYQKEGVLAS
251 GIDGFVYSFI IKDRSYMIED FLEIERPVEH MTFSPNYTVL LIQTDKGSVY
301 IYTFGKEPTL NKVLDACDGK FQAIDFITPG TQYFMTLTYS GEICVWWLED
351 CACVSKIYLN TLATVLACCP SSLSAAVGTE DGSVYFISVY DKESPQVVHK
401 AFLSESSVQH VV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_8p7, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3_8p7, frame 2

Report for DKFZphtes3_8p7.2

[LENGTH]	412
[MW]	46476.62
[pI]	4.91
[KW]	Alpha_Beta

```
SEQ     MATNIPCEVVAFSDRKLKPLIYVYSFPGLTRRTKLKGNILLDYTLLSFSYCGTYLASYSS
PRD     cccccceeeeeeccccceeeeeeccccccccccchhhhhhheeeecccccccccccc

SEQ     LPEFELALWNWESSIILCKKSQPGMDVNQMSFNPMNWRQLCLSSPSTVSVWTIERSNQEH
PRD     cchhhhhhhhcccccceeecccccccccccceeeccccccceeeeecccccceeeeeeeeecchhh

SEQ     CFRARSVKLPLEDGSFFNETDVVFPQSLPKDLIYGPVLPLSAIAGLVGKEAETFRPKDDL
PRD     hhhhhhhccccccccccccccccccccccccccccccceeeeeecccccccccccccccc

SEQ     YPLLHPTMHCWTPTSDLYIGCEEGHLLMINGDTLQVTVLNKIEEESPLEDRRNFISPVTL
PRD     ccccccccccccccceeeecccceeeecccceeeeeehhhhhcccccccccccccccee

SEQ     VYQKEGVLASGIDGFVYSFIIKDRSYMIEDFLEIERPVEHMTFSPNYTVLLIQTDKGSVY
PRD     eeeceeeeecccceeeeeeeeeccchhhhhhhhhhccccceeeccccceeeeeeeccccee

SEQ     IYTFGKEPTLNKVLDACDGKFQAIDFITPGTQYFMTLTYSGEICVWWLEDCACVSKIYLN
PRD     eeeccccccchhhhhccccccceeeeccccceeeeeeeccceeeeeeeecceeeeeeeehh

SEQ     TLATVLACCPSSLSAAVGTEDGSVYFISVYDKESPQVVHKAFLSESSVQHVV
PRD     hhhhhhhcccccceeeecccccceeeeeeeecccccchhhhhhhcccccccc
```

(No Prosite data available for DKFZphtes3_8p7.2)
(No Pfam data available for DKFZphtes3_8p7.2)

DKFZphtes3_9e22
----

group: testes derived

DKFZphtes3_9e22 encodes a novel 227 amino acid protein with weak partial similarity to Ring-finger proteins.

For the novel protein, Pfam, but not Prosite predicts a C3HC4 type RING finger motife.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to zinc finger proteins

Sequenced by DKFZ

Locus: unknown

Insert length: 1318 bp
Poly A stretch at pos. 1308, no polyadenylation signal found

```
   1 GCTCCCCCGG CTTTCGGAGC CCGGGGGCGG CCTGTGGCGC GCGGAGCCCG
  51 CGCCGGACTG CGCCTCTTTG GACCTTGAGG GGAAACATGC GTTTGCCTTG
 101 GATCGTTTGA AATTCTAAGT TTGGGATCCC CGCCCGCCCG CCTGCCTCTT
 151 CCGCCCCGCG GGTTTTTTCC TTTTTTCCTT TTGCTTTTTT TCCTTTTCTC
 201 CCTCCGGGTC TCCTTTTTGA CTCCCTCCCC CTTTATGCTC GCCCAGCCCT
 251 CCCCCTGCTG CTGAGAAGTG GGGGAGGGTC TCGGCCTCCA GGTTCCCGCC
 301 CCACCGGGGC CCGGGCGAGC ATGGGGGGCA AGCAGAGCAC GGCGGCCCGC
 351 TCCCGGGGCC CCTTCCCGGG GGTCTCCACC GATGACAGCG CCGTGCCGCC
 401 GCCGGGAGGG GCGCCCCATT TCGGGCACTA CCGGACGGGC GGCGGGGCCA
 451 TGGGGCTGCG CAGCCGCTCG GTCAGCTCGG TGGCAGGCAT GGGCATGGAC
 501 CCCAGCACGG CCGGGGGGGT GCCCTTTGGC CTCTACACCC CCGCCTCCCG
 551 GGGCACCGGC GACTCCGAGA GGGCGCCCGG CGGCGGAGGG TCTGCGTCCG
 601 ACTCCACCTA TGCCCATGGC AATGGTTACC AGGAGACGGG CGGCGGTCAC
 651 CATAGAGACG GGATGCTGTA CCTGGGCTCC CGAGCCTCGC TGGCGGATGC
 701 TCTACCTCTG CACATCGCAC CCAGGTGGTT CAGCTCGCAT AGTGGTTTCA
 751 AGTGCCCCAT TTGCTCCAAG TCTGTGGCTT CTGACGAGAT GGAAATGCAC
 801 TTTATAATGT GTTTGAGCAA ACCTCGCCTC TCCTACAACG ATGATGTGCT
 851 GACTAAAGAC GCGGGTGAGT GTGTGATCTG CCTGGAGGAG CTGCTGCAGG
 901 GGGACACGAT AGCCAGGCTG CCCTGCCTGT GCATCTATCA CAAAAGCTGC
 951 ATAGACTCGT GGTTTGAAGT GAACAGATCT TGTCCGGAAC ACCCTGCGGA
1001 CTGACCTGCG GGCTTGCTTG CTGACTCCTC TCAAAGGGAC AGAGCGCCCC
1051 TGCTCCAGGG AGGAGGCTCA CCGGACCCTG GGGCAGAGCT GAGCTTGGGA
1101 CACCAGCGGG AACAGGGCAC CCCTTCTGCA CTGACTTCCA GATCATGGTT
1151 CTCCCTTCCT CCCTGAGGAC ACCAAATTGG ATGAGAGCAA GTTTGAGAGA
1201 AGAATGAATC AACTGCTATC CTTCCCCTCA CCCCTCAGCC CAGGAGGGAA
1251 AGGGCATTTT CTTTTTCATC TTTGAAAGGC ATTGTGGGTC TGTCTTTAAA
1301 GTGTTTACAA AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 321 bp to 1001 bp; peptide length: 227
Category: similarity to known protein
Classification: unclassified

```
  1 MGGKQSTAAR SRGPFPGVST DDSAVPPPGG APHFGHYRTG GGAMGLRSRS
 51 VSSVAGMGMD PSTAGGVPFG LYTPASRGTG DSERAPGGGG SASDSTYAHG
101 NGYQETGGGH HRDGMLYLGS RASLADALPL HIAPRWFSSH SGFKCPICSK
151 SVASDEMEMH FIMCLSKPRL SYNDDVLTKD AGECVICLEE LLQGDTIARL
```

```
  201 PCLCIYHKSC IDSWFEVNRS CPEHPAD
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_9e22, frame 3

TREMBL:AF078823_1 product: "RING-H2 finger protein RHA2b"; Arabidopsis thaliana RING-H2 finger protein RHA2b mRNA, complete cds., N = 1, Score = 111, P = 2.8e-06

TREMBL:AF078822_1 product: "RING-H2 finger protein RHA2a"; Arabidopsis thaliana RING-H2 finger protein RHA2a mRNA, complete cds., N = 1, Score = 112, P = 6.6e-06

TREMBL:AC004138_14 gene: "T17M13.17"; Arabidopsis thaliana chromosome II BAC T17M13 genomic sequence, complete sequence., N = 2, Score = 123, P = 1.4e-05

PIR:T02286 hypothetical protein T13D8.23 - Arabidopsis thaliana, N = 1, Score = 142, P = 8.8e-08

```
>PIR:T02286 hypothetical protein T13D8.23 - Arabidopsis thaliana
           Length = 327

  HSPs:

 Score = 142 (21.3 bits), Expect = 8.8e-08, P = 8.8e-08
 Identities = 24/57 (42%), Positives = 30/57 (52%)

Query:   166 SKPRLSYNDDVLTKDAGECVICLEELLQGDTIARLPCLCIYHKSCIDSWFEVNRSCP 222
             S P +      LT D  +C +C+EE + G     LPC  IYHK CI  W  +N SCP
Sbjct:   206 SLPSVKITPQHLTNDMSQCTVCMEEFIVGGDATELPCKHIYHKDCIVPWLRLNNSCP 262
```

Pedant information for DKFZphtes3_9e22, frame 3

Report for DKFZphtes3_9e22.3

```
[LENGTH]        227
[MW]            23782.62
[pI]            6.18
[HOMOL]         PIR:T02286 hypothetical protein T13D8.23 - Arabidopsis thaliana 2e-08
[FUNCAT]        99 unclassified proteins     [S. cerevisiae, YDR313c] 4e-06
[FUNCAT]        30.07 organization of endoplasmatic reticulum       [S. cerevisiae, YOL013c]
0.001
[FUNCAT]        06.13 proteolysis      [S. cerevisiae, YOL013c] 0.001
[PFAM]          Zinc finger, C3HC4 type (RING finger)
[KW]            Irregular

SEQ    MGGKQSTAARSRGPFPGVSTDDSAVPPPGGAPHFGHYRTGGGAMGLRSRSVSSVAGMGMD
PRD    cccccccccccccccccccccccccccccccccccccccccccccccccceeecccccccc

SEQ    PSTAGGVPFGLYTPASRGTGDSERAPGGGGSASDSTYAHGNGYQETGGGHHRDGMLYLGS
PRD    ccccccccccccccccccccccccccccccccccccccccccccccccccccccceeech

SEQ    RASLADALPLHIAPRWFSSHSGFKCPICSKSVASDEMEMHFIMCLSKPRLSYNDDVLTKD
PRD    hhhhhhhhceeeccccccccccccccccccccchhhhhhhhhhhcccccccccccccccc

SEQ    AGECVICLEELLQGDTIARLPCLCIYHKSCIDSWFEVNRSCPEHPAD
PRD    cceeeeeeccccccccccccccceeeeeecccchhhhhhhhcccccccc
```

(No Prosite data available for DKFZphtes3_9e22.3)

Pfam for DKFZphtes3_9e22.3

```
HMM_NAME       Zinc finger, C3HC4 type (RING finger)

HMM               *CPICFcTFQlDyPWPFdePmMlPCgHsFCypCIrrW......CPmC*
                   C IC     L+++    D++  LPC+ ++ ++CI +W       CP+
Query          184 CVIC-----LEELLQGDTIARLPCLCIYHKSCIDSWFEVNRSCPEH     224
```

DKFZphtes3_9i20
----

group: testes derived

DKFZphtes3_9i20 encodes a novel 205 amino acid protein with similarity to human KIAA0336 gene.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="44.1 cR from top of Chr17 linkage group"

Insert length: 2509 bp
Poly A stretch at pos. 2499, polyadenylation signal at pos. 2481

```
   1 CTCGCCGAGA TGACCTGGGC ACCTCTGCGT TGAATCGGCA AATACTGATC
  51 AAGCCGCATT TATTCTGCTC TCAGGAACTC TAAGTCTAGC AGAGAAGATG
 101 AGGCGGTAGA AGTTCATCAA TGGCTTGGCT GGAGGACAAG CAAATTGAGG
 151 ACATTGGCAA CGGAGTGATC AAAATGATAG ATCATGAGGC CTAAAATGAA
 201 TAAGGAAAGA AGAGAAGTGG CAGAGGCTGA GAACAGAAAG AGAGGGTGGA
 251 GGGGCTGTAA ATCTTGAAGA TTAGGGTATA ATATGAGTAT ATGGGTAAGA
 301 ATTGGAAGAA TTGTGTAGGA GGCAGTAGTC AAAAAGTAGA AGCAGTTTGG
 351 AAGAGTAGTT ACAAATATCA AGAGCCAGGT GGCTAAAAGG TGGAGCTATA
 401 GGTCATTGAA GCTCAAGAAA CTGAGTCTCT AGGGCATTGG TTAAGTCATC
 451 TGTCTAGACT TCAAAGTTGT CTAGGATGAT AATTCAGAAG ACTGATCTGT
 501 GCCAAAGTCA CAGGTTTTTC ACGACTGAAA ACAACATAGC AAAATAAGCC
 551 AAGATGTCTG TGGATCCAAT GACCTACGAG GCCCAGTTCT TTGGCTTCAC
 601 GCCACAAACG TGCATGCTTC GGATCTACAT TGCATTTCAA GACTACCTAT
 651 TTGAAGTGAT GCAGGCCGTT GAACAGGTTA TTCTGAAGAA GCTGGATGGC
 701 ATCCCAGACT GTGACATTAG CCCAGTGCAG ATTCGCAAAT GCACAGAGAA
 751 GTTTCTTTGC TTCATGAAAG GACATTTTGA TAACCTTTTT AGCAAAATGG
 801 AGCAACTGTT TTTGCAGCTG ATTTTACGTA TTCCCTCAAA CATCTTGCTT
 851 CCTGAAGATA AATGTAAGGA GACACCTTAT AGTGAGGAAG ATTTTCAGCA
 901 TCTCCAGAAA GAAATTGAAC AGTTACAGGA GAAGTACAAG ACTGAATTAT
 951 GTACTAAGCA GGCCCTTCTT GCAGAATTAG AAGAGCAAAA AATTGTTCAG
1001 GCCAAACTCA AACAGACGTT GACTTTCTTT GATGAGCTTC ATAATGTTGG
1051 CAGAGATCAT GGGACTAGTG ATTTTAGGGA GAGTTTAGTA TCCCTGGTTC
1101 AGAACTCCAG AAAACTACAG AACATTAGAG ACAATGTGGA AAAGGAATCG
1151 AAACGACTGA AAATATCTTA ATTGCTCAGT AGTCAAAAGG AGGAGCCTGT
1201 CAAAAAGTAG AATCATAAGG ACTGTTCAAA CCATAAGGAC TGTTCAAATC
1251 ATACCAGTGA CTGTTCAAAC CAACCATACT TTTTATTAGA TTTGCTTTGT
1301 CAACTCTTTC TTGTATTCTG TGTTTTCCTC TTTTTTGGTC CACTTTGCTG
1351 AGGTATGAAG TGTACTACTT TGAACTAGGC TGAAGCATCT GAGTCTTCTA
1401 ATAAGTGGGA AGGGATCCAA CAAAGAAGCC ATGACCAGTT AAAGATATTT
1451 GCAGAGTTAC ACCTTGGTCA TAAGTCCTTT GTGACCTTGA TTATTTTGGC
1501 TTACTCTTTG GATGAGACCA GACAAGAAAA GGATTAAACG GGTGGCTCCT
1551 TTAATATTAT TATTATTGTT TTTGAGACAA GGTCCCTTTC TGTCACCCAG
1601 GTTAGAGTAG ATTTCAGTGG CACAATCTTG GCTCACTGCA ACCTCTGTGT
1651 CCTGGGCTCA AGTGATCCTC CTGCCTCAGC CTCCCAAGTA GCTAGGACCA
1701 CAGGTGCGTG TCACCATGCT TGGCTAATTT TTTTGCAGAA ACGAGGCCTC
1751 ACTATATTGT CCAGGCTGAG TGGCTCTTTT ATTAACCAGT CATTACACTG
1801 CGGAACAGCC AACATAGAGT ACTTGCTCTC GTCCTGTGAA TTTTCTTTCA
1851 TGAGGGAGTC AATATGTAGT GGAAAGAAGC ATGTAGCAAA AAAGACAACC
1901 TTGATCTTTA ATAAAAAAGA AGTTGGTTTA TTTCCAAAAT AAATCCCCTG
1951 ACAAAAAACC TGGTGATGTT AAGCAATTGA CTGTCTTAGA GTCCAGCAGA
2001 AGACCTTAGA CAAAAAAAGC AGAACCCACT GGAGTAGAAA AGGAAGCATG
2051 TAGCATATAC TCAGTAGTGA AATTTAATTT TACTGACTGT TAGGTATCTA
2101 TGCCAATTTG TTTTCATACT TCAGTTGGTT TTGGAATCTG CCTTATACCT
2151 AATATTTATT TATTCACACT CATAAGCATC AAATATTTAA TGCCCTCAGT
2201 GGGAAATTTG TGTTTAAACT CAATGGAATC TAATATTTCT TTATGTCGTT
2251 AGTCCCTGTA AAATGTTAGG TCACCCAAGG AAAGGGGAGA AATAGCAATG
2301 GTTGTTCCTA AGGTATTGCT TGCCCTCCAT GTCTTCCTAA AGAGCAGAAC
2351 TTGGAGTTTC TCCTTTATGT AGAGAAGAAG TAACTTAGGG TGTATTTGCA
2401 ATGAAATATT CATAGATATT GAAAGCTTGT GTTTACATGA AATATGTTTA
2451 TTATCAAGAA GTCCTTTTTC CAATTCTGTA CATTAAATAT ATGTGTTTTA
2501 AAAAAAAAA
```

BLAST Results

Entry AC004148 from database EMBL:
Homo sapiens chromosome 17, clone HCIT524C5, complete sequence.
Score = 5245, P = 0.0e+00, identities = 1049/1049
3 exons

Entry HS556361 from database EMBL:
human STS TIGR-A003N29.
Score = 1005, P = 1.3e-39, identities = 201/201

Entry HSG043 from database EMBL:
human STS SHGC-36031.
Score = 955, P = 2.8e-37, identities = 205/215

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 554 bp to 1168 bp; peptide length: 205
Category: putative protein
Classification: no clue

```
  1 MSVDPMTYEA QFFGFTPQTC MLRIYIAFQD YLFEVMQAVE QVILKKLDGI
 51 PDCDISPVQI RKCTEKFLCF MKGHFDNLFS KMEQLFLQLI LRIPSNILLP
101 EDKCKETPYS EEDFQHLQKE IEQLQEKYKT ELCTKQALLA ELEEQKIVQA
151 KLKQTLTFFD ELHNVGRDHG TSDFRESLVS LVQNSRKLQN IRDNVEKESK
201 RLKIS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_9i20, frame 2

TREMBLNEW:HSAB2334_1 gene: "KIAA0336"; Human mRNA for KIAA0336 gene, complete cds., N = 1, Score = 107, P = 0.0081

```
>TREMBLNEW:HSAB2334_1 gene: "KIAA0336"; Human mRNA for KIAA0336 gene,
     complete cds.
            Length = 1,583

  HSPs:

 Score = 107 (16.1 bits), Expect = 8.2e-03, P = 8.1e-03
 Identities = 42/140 (30%), Positives = 76/140 (54%)

Query:    65 EKFLCFMKGHFDNLFSKMEQLFLQLILRIPSNILLPEDKCKETPYSEED----FQHLQKE 120
             EK  CF+K H +NL    +EQ    +L  R     ILL +D    ++P  + D      + L+++
Sbjct:   796 EKEKCFIKEH-ENLKPLLEQK--ELRDRRAELILL-KDSLAKSPSVKNDPLSSVKELEEK 851

Query:   121 IEQLQE--KYKTELCTKQALLAELEEQKIVQAKLKQTLTFFDELHNVGRDHGTSDFRESL 178
             IE L++   K K E    K   L+A ++ +K + +  K+T T  +EL ++  +        S+
Sbjct:   852 IENLEKECKEKEEKINKIKLVA-VKAKKELDSSRKETQTVKEELESLRSEK--DQLSASM 908

Query:   179 VSLVQNSRKLQNIRDNVEKESKRLKI 204
               L+Q +    +N+     EK+S++L +
Sbjct:   909 RDLIQGAESYKNLLLEYEKQSEQLDV 934
```

Pedant information for DKFZphtes3_9i20, frame 2

Report for DKFZphtes3_9i20.2

[LENGTH]	205	
[MW]	24140.13	
[pI]	5.51	
[KW]	All_Alpha	
[KW]	COILED_COIL	18.05 %

```
SEQ     MSVDPMTYEAQFFGFTPQTCMLRIYIAFQDYLFEVMQAVEQVILKKLDGIPDCDISPVQI
PRD     cccccchhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccc
COILS   ............................................................

SEQ     RKCTEKFLCFMKGHFDNLFSKMEQLFLQLILRIPSNILLPEDKCKETPYSEEDFQHLQKE
PRD     cccchhhhhhcccccchhhhhhhhhhhhhhhcccceeecccccccccccchhhhhhhhhh
COILS   ..................................................CCCCCCCCCC

SEQ     IEQLQEKYKTELCTKQALLAELEEQKIVQAKLKQTLTFFDELHNVGRDHGTSDFRESLVS
PRD     hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccchhhhhhh
COILS   CCCCCCCCCCCCCCCCCCCCCCCCCCC.................................

SEQ     LVQNSRKLQNIRDNVEKESKRLKIS
PRD     hhcccchhhhhhhhhhhhhhhcccc
COILS   .........................
```

(No Prosite data available for DKFZphtes3_9i20.2)

(No Pfam data available for DKFZphtes3_9i20.2)

DKFZphtes3_9k22
----

group: testes derived

DKFZphtes3_9k22 encodes a novel 304 amino acid protein with partial similarity to X. leavis katanin p80.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C-terminus of katanin p80

Sequenced by DKFZ

Locus: unknown

Insert length: 2676 bp
Poly A stretch at pos. 2665, no polyadenylation signal found

```
   1 CTCTCTAGGC TGCCGGGCGC TGGTCGTCAG CGCCGAGGCT GGGCTGAGGC
  51 GCCGCGGTAC CATGAGGCGC CGGTACTTAA GAGATTATGG CATCAGAAAC
 101 CCACAATGTT AAAAAACGGA ACTTTTGTAA TAAGATTGAG GATCATTTCA
 151 TTGATCTTCC TAGAAAAAAG ATCTCTAATT TCACTAATAA GAACATGAAG
 201 GAGGTTAAGA AATCTCCAAA ACAGTTGGCT GCTTACATAA ATAGAACAGT
 251 TGGACAAACT GTGAAAAGCC CAGATAAACT TCGTAAAGTG ATCTATCGCA
 301 GAAAGAAAGT TCATCATCCC TTTCCAAATC CTTGTTACAG AAAAAAACAG
 351 TCCCCTGGAA GTGGGGGCTG TGACATGGCA AATAAAGAAA ATGAACTGGC
 401 TTGTGCAGGC CACCTGCCTG AAAAATTACA CCATGATAGT CGAACATATT
 451 TGGTTAACTC CAGTGATTCT GGTTCTTCAC AGACAGAAAG CCCATCATCA
 501 AAATATAGTG GGTTTTTTTC TGAGGTTTCT CAGGACCATG AAACAATGGC
 551 CCAAGTTTTG TTCAGCAGGA ATATGAGATT GAATGTAGCT TTAACTTTCT
 601 GGAGAAAGAG AAGTATAAGT GAACTTGTAG CTTATTTGTT GAGGATAGAA
 651 GATCTTGGCG TTGTGGTAGA TTGCCTTCCT GTGCTCACCA ATTGTTTACA
 701 GGAAGAAAAA CAATATATCT CACTTGGCTG CTGTGTTGAC TTGTTGCCTC
 751 TAGTAAAGTC ACTACTTAAA AGCAAATTTG AAGAATATGT TATAGTTGGT
 801 TTAAACTGGC TTCAAGCAGT CATTAAAAGG TGGTGGTCAG AACTATCATC
 851 CAAAACAGAA ATTATAAATG ATGGAAATAT TCAAATTTTA AAACAACAAT
 901 TAAGTGGATT ATGGGAACAG GAAAACCATC TTACTTTGGT TCCAGGATAT
 951 ACTGGTAATA TAGCTAAGGA TGTAGATGCT TATTTATTAC AGTTACATTG
1001 AGAGATTTCA TCTACTAAAG AGCATTTGGT TTTTCAAAAC ATCCCTGAAC
1051 TGTATAATTT ACAAAAAAAA AAGTCTCGTC TGAGAACTGT GAACTGTGGA
1101 AGAAATCAAA ACTATTTTTT CTTTTAAAAA GCCACGTAAT GAAACCACTA
1151 ATGAAATCCC AGCAATCTGC TTCACATTGA AGTGGAAAAA TATCCAAAAG
1201 GAGCAGCTTC AATTTCATTG AGGTGAAAGT GCACTATGAA GATTGTTCAC
1251 CTTTGCTGCA TTTGGGAGTT ATATGGTTAT TTGGTAACAT TAAGAACTAC
1301 TGGATTTTAA TGCAATCCTG CATAAAAATA TAATTTATAC TATGTGAAAA
1351 AATAAGACAG GACTTACCAC TAGGAACCAC CAAGACCAAT CATCATTAAC
1401 TTTTTTAAGA TTGTGTTTTA TTAAAAAAAA AAAACACTTA AATGTGTGCA
1451 GCTATTTTCT TATGTTGAAA AGACTGAAAG TTTAAAACAT GAAAAAAATC
1501 AATATTAAAC ATTTTTTGTT CACACTGAGA TACTGTGTAT GTAAAATGCC
1551 TTAATTATTA ATAAGCCAAT GTGTTATGAT ACCAATATCT GTTTTAAAAA
1601 ACTAAAACCA ACCATGCTTC TGGCATGATA AAATCATGGA ATTAAATCAG
1651 GGGTTTACAT TCTTGTAGAG TGTTCTTGAA ACACTCTCTG CACCATTTTT
1701 AAAACTTGAG AATAGTTTTA GTATCTCTGA TATTTTTTGC CAGAATCATC
1751 ATGTCATGTA TGAATGTGTT ATCCCTATCT AAGGAAAAAG GTGAATATGT
1801 TTTTGTATGA ATGTTTAACT GGAAATGTCC ATGGACTTGG CTAATTTATA
1851 TTTACTTTTT ATTGTACATA GATTTCTAAT ATTTTTCATT CCTGTATCAT
1901 TTAAACTTCC TTCATTTGAG TAAATTCACT AAATATTTCT ATTTTTTTGC
1951 TTTTTTAAAT TCTGATTTTA TATGAATTCT AATTCTTTTT CACTACATAT
2001 GTTTTAAAGA GTTACATACA GTGATTTAGA ATGGTTTACA GTTAATGCTG
2051 ATCTTGTATT TTAAATTCCA ACACTTTGTG TCACTACCTC CTCTAATGGT
2101 TAGTATGATA TGCTAGCAGA CTGTATGAGG TCTTTTTTTA AAATACCACT
2151 TTTAGTGTCA GTGAACCAAA TTCTGGAATG TCTTAACAGC TCTAAATCTT
2201 ACTTGTCTTG AAAATGATTG GGGTTTAATA CCACTGCTGG TGGTTCACAC
2251 ATCATCCCAT CCTTAATATG CCTGACAGGC ATCTGAGCAA AGGTTTTTAG
2301 TAATTGAATT TCTCTGCAGT AGTCCTTCAA GCACTTGAAT GTAAACCTTT
2351 AGCATTTATT CGTTTAATGA CTACTGATAC GAATCTCAAG CAGATTTCTT
2401 GCTCTTAAAA GTTATGTTTC ACTGAGTTCT GGTTTTGTGT AGCTATATTT
2451 TATATAGCTA GATATTCCTC ACAGTGAACA TGAATTGTAA TAATTGGTTA
2501 TTTCCTTAAG TCTTTAGATT ATAATAATTT CAGATTATTG CACGTCTGTG
2551 ATTTGACAGG TGAGTTATTT AAGAGGCCAG TTTTCAGGAC ATGGGAATTT
2601 GAATTGTAAA CCTGTTATCT CTGTGAAACT TTTAACATGA TAAAATATAA
2651 CCTTTCTTTG TGCTTAAAAA AAAAAA
```

BLAST Results

Entry HS541354 from database EMBL:
human STS WI-11840.
Score = 1267, P = 7.1e-50, identities = 271/281

Medline entries

98227670:
Katanin, a microtubule-severing protein, is a novel AAA ATPase that targets to the centrosome using a WD40-containing subunit.

Peptide information for frame 3

ORF from 87 bp to 998 bp; peptide length: 304
Category: similarity to known protein
Classification: unclassified

```
  1 MASETHNVKK RNFCNKIEDH FIDLPRKKIS NFTNKNMKEV KKSPKQLAAY
 51 INRTVGQTVK SPDKLRKVIY RRKKVHHPFP NPCYRKKQSP GSGGCDMANK
101 ENELACAGHL PEKLHHDSRT YLVNSSDSGS SQTESPSSKY SGFFSEVSQD
151 HETMAQVLFS RNMRLNVALT FWRKRSISEL VAYLLRIEDL GVVVDCLPVL
201 TNCLQEEKQY ISLGCCVDLL PLVKSLLKSK FEEYVIVGLN WLQAVIKRWW
251 SELSSKTEII NDGNIQILKQ QLSGLWEQEN HLTLVPGYTG NIAKDVDAYL
301 LQLH
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3_9k22, frame 3

TREMBL:AF056021_1 product: "p80 katanin"; Xenopus laevis p80 katanin mRNA, partial cds., N = 1, Score = 146, P = 1.2e-07

TREMBL:AF052432_1 product: "katanin p80 subunit"; Homo sapiens katanin p80 subunit mRNA, complete cds., N = 1, Score = 150, P = 1.2e-07

TREMBL:AF052433_1 product: "katanin p80 subunit"; Strongylocentrotus purpuratus katanin p80 subunit mRNA, complete cds., N = 2, Score = 146, P = 4.2e-07

```
>TREMBL:AF052432_1 product: "katanin p80 subunit"; Homo sapiens katanin p80
     subunit mRNA, complete cds.
            Length = 655

  HSPs:

 Score = 150 (22.5 bits), Expect = 1.2e-07, P = 1.2e-07
 Identities = 35/105 (33%), Positives = 55/105 (52%)

Query:   145 SEVSQDHETMAQVLFSRNMRLNVALTFWRKRSISELVAYLLRIEDLGVVVDCLPVLTNCL 204
             S++ + H+TM  VL SR+  L+     W     I    V   + I DL VVVD L    N +
Sbjct:   489 SQIRKGHDTMCVVLTSRHKNLDTVRAVWTMGDIKTSVDSAVAINDLSVVVDLL----NIV 544

Query:   205 QEEKQYISLGCCVDLLPLVKSLLKSKFEEYVIVGLNWLQAVIKRW 249
              ++       L  C  +LP ++ LL+SK+E YV  G   L+ +++R+
Sbjct:   545 NQKASLWKLDLCTTVLPQIEKLLQSKYESYVQTGCTSLKLILQRF 589
```

Pedant information for DKFZphtes3_9k22, frame 3

Report for DKFZphtes3_9k22.3

[LENGTH]	304
[MW]	34767.24
[pI]	9.18
[KW]	All_Alpha

```
[KW]        LOW_COMPLEXITY      3.95 %

SEQ     MASETHNVKKRNFCNKIEDHFIDLPRKKISNFTNKNMKEVKKSPKQLAAYINRTVGQTVK
SEG     ............................................................
PRD     ccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhccccc

SEQ     SPDKLRKVIYRRKKVHHPFPNPCYRKKQSPGSGGCDMANKENELACAGHLPEKLHHDSRT
SEG     ............................................................
PRD     ccchhhhhhhhhhhccccccccccccccccccccccccchhhhhhcccccccccccccce

SEQ     YLVNSSDSGSSQTESPSSKYSGFFSEVSQDHETMAQVLFSRNMRLNVALTFWRKRSISEL
SEG     ............................................................
PRD     eeecccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ     VAYLLRIEDLGVVVDCLPVLTNCLQEEKQYISLGCCVDLLPLVKSLLKSKFEEYVIVGLN
SEG     ......................................xxxxxxxxxxxx..........
PRD     hhhhhhhhcceeeeeecccchhhhhhhhceeeccceeeehhhhhhhhhhhheeeeeeehh

SEQ     WLQAVIKRWWSELSSKTEIINDGNIQILKQQLSGLWEQENHLTLVPGYTGNIAKDVDAYL
SEG     ............................................................
PRD     hhhhhhhhhhhhcccceeeecccccccccccccccchhhhhhhhhccccccccchhhhhhh

SEQ     LQLH
SEG     ....
PRD     hccc
```

(No Prosite data available for DKFZphtes3_9k22.3)

(No Pfam data available for DKFZphtes3_9k22.3)

Localization of expressed proteins

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp434B0435	AL117496	6248	48	5366	1773	transport and traffic	similar to: kinesin like proteins	512.1 cR from top of Chr10 linkage group	"secr pathway"	None
DKFZp434N0535	AL117518	4055	126	4025	1300	differentiation & development	similar to: Drosophila chromatin protein		"no predict"	None
DKFZp564A0122	AL110209	2722	65	1300	412	signaling & communication	similar to: acyltransferase	16	"mitochondria"	Mitochondria
DKFZp564A022	AL136620	1376	132	632	167	unknown	unknown	4	"no predict"	Endoplasmic Reticulum
DKFZp564A032	AL50267	2214	76	1953	626	differentiation & development	similar to: MG21 contains three conserved protein motifs present in GTP-binding proteins, but these are not conserved in 2_2a3.1	238.7 cR from top of Chr20 linkage group	"no predict"	Nucleus
DKFZp564A0723	AL80116	2524	42	2177	712	cell cycle	similar to: origin recognition complex	6q14.3-16.1	"nucleus"	Cytosol + Nucleus
DKFZp564A202	AL80056	707	33	620	196	metabolism	similar to: protein involved in energy metabolism	10	"mitochondria"	Endoplasmic Reticulum
DKFZp564B0482	AL110243	2092	317	1579	421	signaling & communication	Unknown, contains 2 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins		"no predict"	Cytosol + Nucleus
DKFZp564B1023	AL136611	2905	157	1896	580	nucleic acid management	similar to: RNA helicase		"nucleus / nuclear envelope"	Nucleus

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp564B1162	AL136646	4593	661	2625	655	signaling & communication	unknown	4	"no predict"	Cytoskeleton
DKFZp564B122	AL049972	1842	70	1536	488	unknown	unknown		"no predict"	Cytosol
DKFZp564B1471	AL136667	1484	78	323	82	membrane protein	unknown		"no predict"	Endoplasmic Reticulum
DKFZp564B162	AL136621	1914	246	1631	462	nucleic acid management	similar to: Zinc finger protein	13q12	"no predict"	other/unknc
DKFZp564B163	AL50268	1208	191	577	129	transport and traffic	similar to: GTP binding protein		"secr pathway"	Cytosol + Nucleus
DKFZp564B212	AL136623	1915	218	1348	377	protein management	similar to: protein involved in posttranslational modification	22q12.1	"secr pathway"	Endoplasmic Reticulum
DKFZp564B2123	AL136612	3300	121	699	193	signaling & communication	similar to: Neurocalcin is a Ca(2+)-binding protein with 3 EF-hands. Homology with recoverin indicates involvement in Ca2+ dependent activation of guanylate cyclase.	574.6 cR from top of Chr8 linkage group	"no predict"	Cytosol + Nucleus
DKFZp564B246	AL136664	2054	73	1074	334	metabolism	similar to: dTDP-6-deoxy-L-mannose-dehydrogenase	5	"nucleus or cytosol"	Nucleus
DKFZp564C0362	AL80076	1731	60	1142	361	nucleic acid management	similar to: ssDNA binding protein		"no predict"	other/unknown
DKFZp564C0469	AL050298	899	86	898	270	unknown	unknown		"nucleus"	Cytosol + Nucleus
DKFZp564C1362	AL136647	837	137	673	179	metabolism	similar to: molecular clock protein	16p12.3-p13.11	"mitochondria"	Other/unknown

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp564C1616	AL136597	3928	240	1997	586	structure & motility	shares the features of mayven and kelch and therefore should be involved in the organisation of cyto skeleton binding to membrane proteins		"cytoskeleton / plasma membrane"	Nucleus
DKFZp564C162	AL136627	2305	155	625	157	membrane protein	unknown	86.2 cR from top of Chr1 linkage group	"no predict"	Endoplasmic Reticulum
DKFZp564C1664	AL136656	1866	180	1040	287	unknown	unknown	745_A_2; 756_F_2; 842_C_2	"no predict"	Cytosol
DKFZp564C182	AL136628	2835	272	1177	302	unknown	unknown		"no predict"	Golgi
DKFZp564C183	AL136639	1709	105	1448	448	nucleic acid management	similar to: DEAD-box helicase	87.50 cR from top of Chr16 linkage group	"nucleus / nuclear envelope"	Nuclear envelope
DKFZp564C196	AL050020	2266	366	966	200	signaling & communication	similar to: neuronal calcium sensor		"no predict"	Nucleus
DKFZp564D116	AL050022	2535	29	1849	607	signaling & communication	similar to: GTP-binding protein		"no predict"	Cytosol
DKFZp564D202	AL136631	1787	18	944	309	unknown	unknown		"no predict"	Cytosol
DKFZp564E0123	AL136613	2005	104	1000	299	unknown	unknown	16q13	"no predict"	Cytosol + Nucleus
DKFZp564E0482	AL136697	2923	163	1581	473	signaling & communication	similar to: calmodulin-related protein	200.5 cR from top of Chr3 linkage group	"membranes"	Cytosol + Nucleus

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp564E1782	AL136696	1618	40	972	311	membrane protein	unknown	171.7 cR from top of Chr14 linkage group	"no predict"	Endoplasmic Reticulum
DKFZp564E2182	AL50261	2367	193	804	204	Cell Cycle	similar to: protein involved in cell cycle, DNA repair, maintenance of minichromosomes	6q22.1-22.33	"nucleus"	Nucleus
DKFZp564F0223	AL136614	1016	68	613	182	unknown	unknown	12q24	"secr pathway / endosomes"	other/unknown
DKFZp564F052	AL049989	1649	34	1303	423	signaling & communication	similar to: sorting nexin 7		"membranes"	Cytosol
DKFZp564F0522	AL049943	2078	283	943	220	unknown	unknown	2	"no predict"	Nucleus
DKFZp564F1862	AL80081	1987	250	918	223	differentiation & development	similar to: DnaJ proteins, but lacks CRR domain of these proteins.	7q31	"no predict"	Endoplasmic Reticulum
DKFZp564F2116	AL136598	1512	115	738	208	membrane protein	unknown	15q25	"nucleus"	other/unknown
DKFZp564F2122	AL136604	1910	156	1856	567	unknown	unknown	311.4 cR from top of Chr14 linkage group	"no predict"	Cytoskeleton (microtubules)
DKFZp564F2162	AL136648	1549	95	730	212	unknown	unknown	209.8 cR from top of Chr20 linkage group	"peroxisomes"	Peroxisomes
DKFZp564G0222	AL80115	1165	157	933	259	nucleic acid management	unknown		"no predict"	Endoplasmic Reticulum

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
DKFZp564G083	AL136641	1027	37	570	178	protein management	similar to: yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1.	20	"no predict"	Cytosol + Nucleus
DKFZp564G182	AL136632	2444	539	1225	229	unknown	unknown	6p22.1-22	"no predict"	Cytosol + Nucleus
DKFZp564H012	AL136633	957	93	632	180	unknown	unknown		"no predict"	Mitochondri
DKFZp564H1122	AL136605	1734	159	1133	325	membrane protein	unknown	11q14	"no predict"	Nucleus
DKFZp564H1322	AL136606	2292	270	1829	520	membrane protein	unknown	19q13.2 from BCKDHA-D19S217	"no predict"	Cytosol + Nucleus
DKFZp564H1562	AL136649	2014	75	971	299	structure & motility	similar to: Cell cell interaction protein	1	"plasma membrane"	Plasma membrane + cell contact sites
DKFZp564I0123	AL136615	1467	126	1064	313	signaling & communication	similar to: protein activator of the interferon-induced protein kinase		"cytosol or nucleus"	Cytosol
DKFZp564I0422	AL136607	4748	511	1194	228	signaling & communication	unknown		"no predict"	Golgi + Plasma membrane
DKFZp564I1216	AL136600	1548	81	635	185	membrane protein	unknown	873.3-875.1 cR from top of Chr1 linkage group	"no predict"	Endoplasmic Reticulum
DKFZp564I1782	AL136699	1741	168	410	81	signaling & communication	similar to: phospholemman protein,	11q23	"secr pathway"	Golgi + plasma

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
							a membrane substrate for the cAMP-dependent protein kinase; seems to serve as chloride channels or as chloride-channel regulators. Transmembrane Protein			membrane
DKFZp564I206	AL136665	1122	34	921	296	unknown	unknown	377.5 cR from top of Chr8 linkage group	"mitochondria"	Mitochondrí
DKFZp564I2423	AL136616	1713	58	882	275	metabolism	similar to: protein involved in amino acid metabolism	8p11.2	"cytosol"	Cytosol + Nucleus
DKFZp564I2482	AL136700	1860	10	1650	547	nucleic acid management	similar to: Dead-box helicase	175.5 cR from topFT of Chr7 linkage group	"nucleus / nuclear envelope"	Nucleus
DKFZp564J1022	AL110301	1409	5	1021	290	nucleic acid management	Unknown, contains a Leucine zipper	12	"cytosol or nucleus"	Cytosol + Nucleus
DKFZp564J1516	AL136601	2868	352	1839	496	structure & motility	similar to: RNA bindng, Tubulin binding	20, 12.10 cR from GCT10F11	"cytosol"	Cytosol
DKFZp564J1864	AL136660	690	109	648	180	transport and traffic	similar to: canin and chicken microsomal signal peptidase 23 kd subunit.		"endoplasmic reticulum"	Endoplasmic Reticulum
DKFZp564J2222	AL136608	1858	154	1440	429	structure & motility	similar to: actin-related protein		"plasma membrane / cytoskeleton"	Plasma membrane
DKFZp564K0322	AL136609	2775	779	2392	538	unknown	unknown		"no	Plasma

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
									predict"	membrane
DKFZp564K0822	AL136610	2789	10	525	172	unknown	unknown	7	"no predict"	Golgi
DKFZp564K1216	AL49933	1938	357	1418	354	signaling & communication	similar to: GTP-binding regulatory protein	7	"membranes"	Golgi + Plasma membrane
DKFZp564K192	AL136637	1931	107	1015	303	unknown	unknown	6p22.1-22.3	"no predict"	Other/unknown
DKFZp564K1964	AL117619	1560	207	884	226	unknown	unknown	17	"no predict"	Endoplasmic Reticulum
DKFZp564K2216	AL136602	2088	832	1155	108	unknown	unknown		"no predict"	Mitochondria
DKFZp564L023	AL136643	2978	279	2045	589	protein management	Unknown, Pfam prediction: ubiquitin family	9	"cytosol"	Cytosol + Nucleus
DKFZp564L1216	AL136603	2042	73	873	267	membrane protein	unknown		"secr pathway"	Golgi + plasma membrane
DKFZp564L2423	AL136617	2416	29	1072	348	transport and traffic	Unknown, a lectin character is predicted	2	"endoplasmic reticulum"	Endoplasmic Reticulum
DKFZp564M082	AL80071	902	227	589	121	unknown	Unknown, contains osteopontin motive		"no predict"	Cytosol + Nucleus
DKFZp564M112	AL80070	2686	14	595	194	signaling & communication	unknown	956.7 cR from top of Chr2 linkage group	"no predict"	Golgi
DKFZp564M173	AL136644	636	26	400	125	unknown	similar to: janus proteins		"no predict"	Cytosol + Nucleus
DKFZp564M1863	AL117602	1192	125	1027	301	signaling &	similar to: phosducin-like protein, G-protein	9	"cytosol"	Cytosol

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
						communication	modulator			
DKFZp564M1982	AL390217	2707	302	1160	286	unknown	unknown		"no predict"	Cytosol
DKFZp564M2423	AL80119	2201	86	1246	387	unknown	unknown	72.60 cR from top of Chr3 linkage group	"cytosol"	Cytosol
DKFZp564N0582	AL50264	1646	75	506	144	cell cycle	similar to: DRR1 gene	3p21.1	"cytoskeleton / plasma membrane"	Cytoskeletc (focal adhesion sites) + nucleus
DKFZp564N1623	AL136618	2936	172	1047	292	signaling & communication	Unknown, contains a WW domain which binds proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes		"no predict"	Cytosol + Nucleus
DKFZp564O043	AL050390	2515	186	1509	441	structure & motility	similar to: ankyrin	7	"no predict"	Cytosol + Nucleus
DKFZp564O0523	AL136619	1736	24	1103	360	unknown	unknown	7q21-q22	"no predict"	Nucleus
DKFZp564O123	AL80122	1985	234	872	213	unknown	unknown		"no predict"	Cytosol + Nucleus
DKFZp564O1762	AL136652	1260	56	901	282	signaling & communication	similar to: low-density lipoprotein (LDL) receptors are the major		"secr pathway"	Golgi

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
							cholesterol-carrying lipoproteins of plasma. The novel protein contains an additional leucine zipper suitable for protein-protein interaction.			
DKFZp564O1923	AL050295	2091	237	2090	617	metabolism	similar to: dTDP-6-deoxy-L-mannose-dehydrogenase		"secr pathway"	Cytosol
DKFZp564O2423	AL390214	3564	656	1072	139	unknown	Unknown, contains CAAX box (prenyl group binding site); found in :Ras proteins, and ras-like proteins such as Rho, Rab, Rac, Ral, and Rap; nuclear lamins A and B; Some G protein alpha subunits, G protein gamma subunits; some dnaJ-like proteins		"no predict"	Cytosol + Nucleus
DKFZp564O243	AL050015	1074	23	834	270	unknown	unknown	3	"no predict"	Endoplasmic Reticulum
DKFZp566I1024	AL050037	1783	5	970	322	unknown	similar to: hypothetical protein Rv0712 - Mycobacterium tuberculosis		"no predict"	Cytosol
DKFZp566J2046	AL136720	1706	16	678	221	metabolism	similar to: 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.	16	"no predict"	Mitochondria
DKFZp566K144	AL136727	3084	456	1079	208	transport and traffic	similar to: Rab protein		"secr pathway"	Golgi
DKFZp586D0919	AL050100	2777	48	494	148	unknown	unknown	12	"no	Golgi

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
									predict"	
DKFZp586E1124	AL136942	2005	184	861	226	transport and traffic	similar to: golgi transmembrane spanning transporter	8	"Golgi"	Golgi + plasma membrane
DKFZp586E1323	AL136936	1854	367	954	196	protein management	similar to: heat shock protein	578.9 cR from top of Chr12 linkage group	"cytosol or nucleus"	Cytosol + Nucleus
DKFZp586E1519	AL050101	2140	82	1680	559	unknown	similar to: A.thaliana A_IG002N01		"no predict"	Cytosol
DKFZp586F1918	AL050091	3489	184	594	137	unknown	unknown		"no predict"	Cytosol + Nucleus
DKFZp586F1919	AL136915	2024	134	745	204	membrane protein	unknown	14.8 cR from top of Chr20 linkage group	"secr pathway"	Golgi + plasma membrane
DKFZp586H2219	AL50282	1971	199	1623	475	unknown	unknown	22q11.2-qter	"no predict"	Cytosol
DKFZp586I0418	AL136912	1568	163	822	220	unknown	unknown	7q31	"no predict"	Cytosol + Nucleus
DKFZp586I1520	AL050149	2439	11	1711	566	transport and traffic	similar to: nuclear RanGTP binding protein		"nucleus"	Nucleus
DKFZp586J1023	AL136938	1048	72	749	226	protein management	similar to: glutahione S-transferase / posttranslational modification		"no predict"	Cytosol + Nucleus
DKFZp586J1119	AL136919	2343	28	2151	708	signaling & communication	unknown		"membranes"	Endoplasmic Reticulum
DKFZp586J1923	AL050220	745	49	588	179	differentiation &	similar to: serine protease	19	"secr pathway"	Endoplasmic Reticulum

CloneID	AccNo	Contig (bp)	ORFStart (bp)	ORFStop (bp)	ORFSize (aa)	ProteinGroup	Similarity	ChromLocation STS	Localization Predicted	Localization
						development				
DKFZp586K0919	AL50283	1782	204	1316	371	unknown	unknown		"no predict"	Cytosol + Nucleus
DKFZp586L0118	AL136913	1076	45	596	184	protein management	similar to: mitochondrianl Ribosomal S40 protein		"nucleus"	Nucleus
DKFZp586M2420	AL136927	1986	23	1855	611	transport and traffic	similar to: mannosyltransferase	11	"secr pathway"	Endoplasmic Reticulum
DKFZp727E151	AL390215	1957	340	1701	454	membrane protein	similar to: transporter proteins (contains 9 transmembrane domains)		"no predict"	Endoplasmic Reticulum
DKFZp727M111	AL117479	2275	79	1899	633	unknown	unknown		"no predict"	Cytosol
DKFZp727M231	AL117480	2428	56	1681	542	unknown	unknown		"no predict"	Cytosol
DKFZp761G05121	AL118986	4592	107	3613	1169	protein management	similar to: SH3 BINDING PROTEIN		"cytosol"	Cytosol
DKFZp761G18121	AL136548	4117	107	1438	444	signaling & communication	similar to: ALLOGRAFT INFLAMMATORY FACTOR		"nucleus"	Nucleus
DKFZp761I12121	AL136549	4130	139	3894	1252	cell cycle	similar to: p53 inducible protein	5q34	"no predict"	Cytosol
DKFZp761M02121	AL136551	3328	178	2163	662	cell cycle	similar to: p53 regulated PA26-T2 nuclear protein		"nucleus"	Cytosol
DKFZp761O15121	AL136552	4293	112	2421	770	signaling & communication	similar to: semaphorin W	328.8 cR from top of Chr2 linkage group	"secr pathway"	Endoplasmic Reticulum

Table of cDNA clones and related data

Group: cell cycle

CloneID DKFZp	Homology	Function	Group
hfbr2_16g18	Similarity to KIAA0797 and yeast Smt4p	Novel protein with similarities to S. pombe SPAC17A5.07c and the S. cerevisiae Smt4p suppressor of MIF2 gene.; involved in centromer organisation	Cell cycle
hfbr2_2k14	Strong similarity to human N33 tumour suppressor gene	New tumour suppressor gene	Cell cycle
htes3_35b4	Human M-phase phosphoprotein-1	The novel protein is C-terminal identical to human M-phase phosphoprotein-1, which is expressed and phosphorylated in the metaphase. Therefore the novel protein seems to be involved in the mitotic spindle during cell division.	Cell cycle
htes3_35p22	Strong similarity to oncogene 1 (tre-2 locus)	Oncogene	Cell cycle
htes3_7j3	Related to the C-TAK1 Cdc25C associated protein kinase	Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five	Cell cycle
htes3_7p10	Strong similarity to XPMC2 protein	XPMC2 of xenopus rescues several different yeast mitotic catastrophe mutants defective in Wee1/Mik1 kinase function.	Cell cycle
hute1_20m11	Similarity to suppressor protein sds22	Suppressor regulator of protein phosphatase-1	Cell cycle

Group cell structure and motility

Clone ID DKFZp	Homology	Function	Group
hfbr2_16c16	Similarity to Drosophila kelch	Shares the features of mayven and kelch and therefore should be involved in the organisation of cyto skeleton binding to membrane proteins	Structure and motility
hfbr2_2b5	Similarity to collagen proteins	New collagen alpha chain	Structure and motility
htes3_15i5	Strong similarity to "radial spokehead" proteins	Part of sperm motor	Structure and motility
htes3_18l7	Similarity to ankyrins	Putative ankyrin	Structure and motility
htes3_1k11	Strong similarity to mouse ENC-1	Nuclear matrix protein	Structure and motility
htes3_72k15	Strong similarity to Rattus norvegicus actin-filament binding protein Frabin.	FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin binding protein. Modulation of cell structure and motility as well as modulation of the JNK/SAPK pathway.	Structure and motility
htes3_7b22	Similarity to paramyosins	Protein iviolved in motility	Structure and motility
hute1_19g22	Strong similarity to tuftelin/enamelin	New connective tissue protein	Structure and motility
hute1_24j6	Strong similarity Rattus norvegicus cell adhesion regulator (CAR1) mRNA	Cell adhesion regulator (signal transduction molecule influencing cell adhesion to collagen)	Structure and motility

Group Differentiation/Development

CloneID DKFZp...	Homology	Function	Group
hfbr2_2d15	Mus musculus testis-specific Y-encoded-like protein (Tspyl1).	TSPY is believed to function in early spermatogenesis and is a candidate for GBY, the putative gonadoblastoma-inducing gene on the Y-chromosome	Differentiation/Development
htes3_35e21	Similarity to interleukin-7 precursor	New interleukin	Differentiation/Development
hute1_2h3	Strong similarity to mouse E25 and gallus E3-16	Homolog is marker for chondro-osteogenic differentiation	Differentiation/Development

Group kidney derived

CloneID DKFZp	Homology	Function	Group
hfkd2_1j9	Strong similarity to XLCL2 protein, African clawed frog	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfkd2_24e23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfkd2_46a6	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfkd2_46b10	Similarity to C.elegans F25B5.3	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfkd2_46d13	Weak similarity to KE03 protein	Contains a RGD site; No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfkd2_4b6	Similarity to Homo sapiens clone 25003 partial CDS.	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived
hfkd2_4c8	Similarity to KIAA0549 and HAP1 (Huntingtin-associated protein-1)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Kidney derived

Group mammary carcinoma derived

CloneID DKFZp.n.	Homology	Function	Group
hmcf1_1c23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Mammary Carcinoma derived
hmcf1_1g13	Similarity to KIAA0766; very weak similarioty to transposases	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Mammary Carcinoma derived

Group Nucleic acid management

CloneID DKFZp...	Homology	Function	Group
hfbr2_23b10	Similarity to rat RNA helicase HEL117	RNA helicase	Nucleic Acid Management
hfbr2_3c18	Strong similarity to RNA helicase and RNA-dependent ATPase from the DEAD box family	DEAD-box	Nucleic Acid Management
hfbr2_64a15	Similarity to inorganic pyrophosphatases (unspliced)	Inorganic pyrophosphatase	Nucleic Acid Management
hfbr2_6o17	Strong similar to RNA helicases	RNA helicases	Nucleic Acid Management
hfbr2_72b18	Similarity to DNA damage induced genes	Similar to dinP of E. coli, yqjH of B. subtilis, dinP of M. tuberculosis and T19K24.15 of A. thaliana. The dinB/P pathway is a second SOS-pathway in E. coli	Nucleic Acid Management
hfbr2_72l12	Similarity to YDR126w	DNA binding protein	Nucleic acid managment
hfbr2_82i24	Strong similarity to DEAD-box subfamily ATP-dependent helicase	Dead-box helicase	Nucleic Acid Management
htes3_14h21	Strong similarity to RNA helicases	RNA helicase	Nucleic Acid Management
htes3_15j3	Similarity to YGR276c, a ribonuclease H of S. cerevisiae.	Rnase H	Nucleic Acid Management
htes3_20m18	Similarity to the S. cerevisiae mitochondrial carrier protein RIM2.	The novel protein contains a leucine zipper and a Prosite mitochondrial energy transfer proteins signature. It is member of a family of substrate carrier proteins which are found in the inner mitochondrial membrane and are involved in energy transfer.	Nucleic Acid Management
htes3_22g2	KIAA0829 is shorter, nearly identical to rat TIP120	Involved in TATA box binding complex	Nucleic Acid Management
htes3_2m18	Nearly identical to mouse Dhm1	Multifunctional nuclease/exoribonuclease	Nucleic acid management
htes3_7p9	Similarity to nuclear domain 10 protein NDP52	Transcription control	Nucleic Acid Management
htes3_8m10	Strong similarity to polyadenylate-binding proteins.	The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA.	Nucleic Acid Management
hute1_18l1	Strong similarity to S.cerevisiae YHR148w	Mitochondrianl Ribosomal S40 protein	Nucleic Acid Management

Group testis associated

CloneID DKFZp...	Homology	Function	Group
htes3_14g5	Strong similarity to cell growth regulating nucleolar protein LYAR, of mouse	Contains a ATP/GTP-binding site motif A (P-loop), but not the zinc finger motif and and nuclear localization signals of lyar. No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_14p14	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_14p7	Weak similarity to kinesin associated protein KAP3	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15a13	Similarity to S.cerevisiae Hop1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15g14	Similarity to YOR243c	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15h1	Weak similarity to Hsp70/Hsp90 organizing protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_15j18	Unknown	Unknown; no predictive prosite pfam or SCOP motife	Testes associated
htes3_17f10	T23E7.2B PROTEIN	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_18f3	Similarity to TNF-inducible protein CG12-1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_19f19	Weak similarity to S. cerevisiae protein YFL046w.	The protein contains a RGD cell attachment site. No informative BLAST results; No predictive prosite, pfam or SCOP motife.	Testes associated
htes3_19j17	Partial similarity to C.elegans Y40B1A.2 protein.	The novel protein contains two Prosite WW/rsp5/WWP domain signatures. No informative BLAST results; No predictive prosite, pfam or SCOP motife.	Testes associated
htes3_20c21	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motife.	Testes associated
htes3_21n23	Strong similarity to rat 7acomp protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_22c23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_22n13	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_27o14	Similarity to C.elegans C55A6.1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_28d14	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2a11	Similarity to mucin	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2d15	Similarity to C.elegans F25H2.1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2f14	Weak similarity to omega protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2g7	Similarity to neurofilament proteins	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2h15	Similarity to S.pombe cdc23	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_2l19	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motife.	Testes associated

Clone ID DKFZp	Homology	Function	Group
htes3_2m20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motife.	Testes associated
htes3_2n9	Very weak similarity to Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23.	No informative BLAST results; No predictive prosite, pfam or SCOP motife.	Testes associated
htes3_30f4	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motife.	Testes associated
htes3_35g6	Strong similarity to R27216_1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_35n24	Unknown	Contains an IG_MHC pattern No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_35p17	Similarity to S.cerevisiae VAC8 and beta-Catenin, but contains no amadillo motifes	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_4b4	Rattus norvegicus late gestation lung protein 1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_4f17	Similarity to KIAA0333 Methyl-CpG binding protein; does not contain such a motife.	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_4o19	Similarity to mucin	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_50j4	Unknown, prolin rich protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_50n23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_50n6	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_6b21	Similarity to KIAA0256	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_6d16	WUGSC:H_DJ1185I07.2, differences to genmodel	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_72k11	Similarity to S.pombe hypothetical repeat-containing protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_7d17	Similarity to KIAA0454	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_7j8	WUGSC:H_DJ1159O04.1 similarity to YBL104p	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_8g11	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_8g5	KIAA087, alternative spliced	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_8p7	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_9e22	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_9i20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated
htes3_9k22	Similarity to C-terminus of katanin p80	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Testes associated

Group transmembrane proteins

CloneID DKFZp	Homology	Function	Group
hfbr2_16i12	Similarity to Fugu rubripes PUT2	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motife.	Transmembrane protein
hfbr2_16l12	Similarity to gallus putative transmembranee protein E3-16	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hfbr2_22h13	Similarity to Drosophila melanogaster EG:39E1.3.	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hfbr2_2b17	Similarity to Drosophila hypothetical 30K protein	3 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hfbr2_2d17	Unknown	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hfbr2_64k24	Similarity to several proteins	5 transmembrane regions. No informative BLAST results; No predictive prosite, pfam or SCOP motife.	Transmembrane protein
hfbr2_82c20	Similarity to C.elegans D1007.5	7 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hfbr2_82e17	Similarity to C.elegans "R01B10.5"	6 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hfbr2_82g14	Unknown proline rich protein	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hfkd2_24a15	Similarity to C. elegans R07G3.8	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hfkd2_3i13	Similarity to A.thaliana YUP8H12.2	3 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hfkd2_4m11	Weak similarity to YMR034c	4 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hmcf1_1a11	Similarity to YDR255c and SPBC29A3.03c	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hmcf1_1e15	Similarity to D-XYLOSE TRANSPORTER	Transporter; 9 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
htes3_15c6	Unknown	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
htes3_2o13	Partial similarity to the IL-17 receptor.	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
htes3_27k4	Strong similarity to C.elegans K07H8.2/ZK185.2	Contains a leucine zipper 10 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive;	Transmembrane protein
htes3_2h1	Similarity to C.elegans C13F10.5	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
htes3_35k24	Unknown	5 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hute1_19f19	Similarity to mouse P24 protein	2 transmembrane domains No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein
hute1_24c19	Unknown	1 transmembrane domain No informative BLAST results; No predictive prosite, pfam or SCOP motive	Transmembrane protein

Group Brain derived

CloneID DKFZp	Homology	Function	Group
hfbr2_16f21	Strong similarity to zinc finger protein 216 has no zn finger, is only similar	PROSITE: Contains no Zinc finger; No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_16k22	Weak similarity to thioredoxin	No informative BLAST results; No predictive prosite, pfam or SCOP motife	Brain derived
hfbr2_22f21	Weak similarity to C.elegans C18C4.5	No informative BLAST results; No predictive prosite, pfam or SCOP motife	Brain derived
hfbr2_22i4	Similarity to Human P52rIPK N-terminus	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_22k3	Weak homology with : EXTENSIN (PROLINE-RICH GLYCOPROTEIN)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_22k8	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motife	Brain derived
hfbr2_23f2	Similarity to Vps29p; saccharomyces cerevisiae (baker's yeast) pep11 protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_23o24	Similarity to CAAX-box protein	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_23o5	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motife	Brain derived
hfbr2_2a2	Similarity to 52K autoantigen Ro/SS-A - human	No informative BLAST results; No predictive prosite, pfam or SCOP motife	Brain derived
hfbr2_2c1	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_2c18	Weak similarity to cyclin-dependent kinase p130-PITSLRE	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_2d20	Similarity to Synechocystis sp. (PCC 6803)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_2g18	J30M3.2 extension of genmodel	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_2h1	Similarity to C.elegans D2007.4 protein	No informative BLAST results; No predictive prosite, pfam or SCOP motife	Brain derived
hfbr2_2h10	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motife	Brain derived
hfbr2_2k19	Similarity to KIAA0378	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_3f16	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_3l2	Weak similarity to ubiquitin-like protein DSK2 yeast	Pfam: ubiquitin family; No informative BLAST results; No predictive prosite or SCOP motive	Brain derived
hfbr2_62n10	Similarity to reticulocyte-binding protein	Contains a Leucine zipper; No informative BLAST results; No predictive pfam or SCOP motive	Brain derived
hfbr2_64a11	Similarity to Drosophila irregular chiasm C-roughest precursor (frame shift)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived

Clone ID DKFZp	Homology	Function	Group
hfbr2_64c16	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_64c4	Similarity to A. thaliana T08I13.5	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_64h6	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_64i20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_64o16	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_6a17	Weak similarity to finger protein zfOC1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_6i20	Similarity to ribosomal protein L15 precursor, mitochondrial	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_71o20	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_72d13	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_72m16	Similarity to C.elegans H14A12.3	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_72n12	Strong similarity to rat Ganglioside expression factor (GEF-2) but even higher identity with C.elegans putative protein Identities = 91/116 (78%)	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_78d13	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_78n23	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_7a24	DKFZphfbr2_7a24.1 similarity to C-terminus of TGF-beta-activated kinase	Only c-terminus homolog; contains no kinase domain; No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_7e22	Similarity to cytochrome b561	No heme domain but a c may helix loop helix signature No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_7j4	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived
hfbr2_82m16	Very weak similarity to A.thaliana F28A23.140	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Brain derived

Group Intracellular Transport and Trafficking

CloneID DKFZp	Homology	Function	Group
hfbr2_23124	Strong similarity to human GP36b glycoprotein and canine VIP 36	A lectin character is predicted. Due to the intracellular localisation of the homologe proteins, it should be involved in cell trafficing	Transport and traffic
hfbr2_2i17	Strong similarity to rab1	GTP binding	Transport and traffic
hfbr2_41m15	Strong similarity to ras-related GTP-binding protein Rab17	GTP-binding, signal transduction	Transport and traffic
hfbr2_62f10	Strong similarity to zinc transporter proteins	Zinc transporter protein	Transport and traffic
hfbr2_62l19	2 nearly identical to dog GTP-binding protein rab10	GTP binding	Transport and traffic
hfbr2_64j18	Strong similarity to dog signal peptidase (EC 3.4.99.-)	Identical to canin and chicken microsomal signal peptidase 23 kd subunit.	Transport and traffic
hfkd2_24n20	Strong similarity to eps8 binding protein e3B1	Contains an Src homology domain 3 and is similar to human eps8 SH3 domain binding protein 1 (e3B1) and spectrins. The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton	Transport and traffic
hfkd2_24p5	Human ankyrin G (ANK-3) new splice variant	New ankyrin protein	Transport and traffic
hfkd2_4k14	Strong similarity to Rab6	New Rab protein	Transport and traffic
htes3_1g13	Similarity to 256 kD golgin, strong similarity to rat "cp151"	New golgin protein	Transport and traffic
htes3_17n18	TonB-dependent receptor protein signature 1	Involved in receptor-mediated uptake	Transport and traffic
htes3_21l16	Identical to rat ribosome attached membrane protein 4	Responsible for transport of proteins into ER	Transport and traffic
htes3_23l11	Nearly identical to mouse ADP-ribosylation-like factor homolog 6 (Arl6).	Protein secretion through the endoplasmic reticulum and the Golgi vesicular trafficking system is initiated by the binding of ADP-ribosylation factors	Transport and traffic
htes3_26g22	Similarity to kinesins.	The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a kinesin motor domain signature. Kinesin is a microtubule-associated force-producing protein that play a role in organelle transport.	Transport and traffic
htes3_4h6	Strong similarity to Kinesin light chain	New kinesin light chain	Transport and traffic
htes3_72p16	Strong similarity to mouse MEM3 and yeast VPS35	New vacuolar protein sorting-associated protein	Transport and traffic
hutel_19h17	Strong similarity to C.elegans ZK1086.1	Steroid turnover in cells	Transport and traffic
hutel_20h13	Strong similarity to alpha-adaptins	New adaptin chain (clathrin assembly protein complex 2 alpha-a large chain)	Transport and traffic
hutel_24e11	Similarity to golgi 4-transmembrane spanning transporter mtp	New golgi transmembrane spanning transporter	Transport and traffic

Group signal transduction

CloneID DKFZp	Homology	Function	Group
hfbr2_23b21	Nearly identical to bovine neurocalcin	Neurocalcin is a Ca(2+)-binding protein with 3 EF-hands. Homology with recoverin indicates involvement in Ca2+ dependent activation of guanylate cyclase.	Signal transduction
hfbr2_23n16	Similarity to putative phosphatidylinositol-4-phosphate 5-kinase	Contains a WW domain which binds proteins with particular proline- motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes	Signal transduction
hfbr2_2c17	(similarity to YMR131c and retinoblastoma-binding protein RbAp46)	The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins.	signal transduction
hfbr2_62b11	Putative GTPase-activating protein, related to human chimaerins	The new protein is expected to activate p21rac-related small GTPases	Signal transduction
hfbr2_78c24	Strong similarity to guanylate-binding proteins (GBPs)	Modulating/blocking the response of cells to interfersons.	Signal transduction
hfbr2_82e4	Strong similarity to rat calmodulin-binding protein	Involved in calmodulin-related pathway	Signal transduction
hfbr2_82i17	Similarity to plasma membrane substrate for cAMP-dependent protein kinase	Transmembrane protein with strong similarity to the phospholemman protein, a membrane substrate for the cAMP-dependent protein kinase; seems to serve as chloride channels or as chloride-channel regulators. Transmembrane Protein	Signal Transduction
hfbr2_82m6	Strong similarity to mouse "sphingosine kinase	Sphingosine kinase	Signal transduction
hfkd2_46m4	Nearly identical to mouse GTP-binding protein	GTP-binding protein	Signal transduction
htes3_15k11	KIAA0781, 5' extension	Heart development/signal transduction	Signal transduction
htes3_1c1	Similarity to GTPase-activating proteins	GTPase-activating proteins	Signal transduction
htes3_1n3	Similarity to Tup1p	Beta-transducin subunit of G-proteins	Signal transduction
htes3_20k2	Strong similarity to rat vanilloid receptor subtype 1.	VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. The novel protein is the human orthologue of rat VR1.	Signal transduction
htes3_21d4	Similarity to RCC1-like G exchanging factor RLG	RCC1 is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein.	Signal transduction
htes3_23n19	Similarity to rat protein kinase C-interacting RBCC protein 1	Protein kinase C-interacting protein	Signal transduction
htes3_4f5	Similarity to S.pombe "beta-transducin"	Contains 3 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins; in addition, a Cytochrome C family heme-binding site signature is present.	Signal Transduction
htes3_6c11	Strong similarity to YNL132w	Could be a steroid receptor	Signal transduction
htes3_8e24	Related to yeast YGL099w and mouse MMR1 putative GTP-binding proteins.	a novel 658 amino acid putative GTP-binding protein	Signal Transduction
hute1_20g21	Ras inhibitor	Receptor tyrosine kinase (RTK)/RAS/MAP kinase signaling cascade	Signal transduction
hute1_22d2	Similarity to GTP-binding proteins	GTP-binding proteins	Signal transduction

hute1_22e12	Strong similarity to S.cerevisiae YGL054c and cornichon	The Drosophila cni and mammalian proteins cornicon are part of a signal transduction pathway involving hte EGF-receptor	Signal transduction

Group Metabolism

CloneID DKFZp	Homology	Function	Group
hfbr2_3g8	Similarity to N-terminal Acetyltransferase Complex ARD1 homolog	In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1.	Metabolism
hfbr2_62o17	Similarity to apolipoprotein E receptor	Low-density lipoprotein (LDL) receptors are the major cholesterol-carrying lipoproteins of plasma. The novel protein contains an additional leucine zipper suitable for protein-protein interaction.	Metabolism
hfbr2_6b24	Similar to dTDP-6-deoxy-L-mannose-dehydrogenases	DTDP-6-deoxy-L-mannose-dehydrogenase	Metabolism
hfbr2_78k24	Similarity to Mus musculus ubiquitin specific protease UBP43.	The novel protein contains a Prosite ubiquitin carboxyl-terminal hydrolases family 2 signature 2. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquinated protein	Metabolism
hfkd2_24b15	Similarity to phosphomannomutases	Phosphoserine signature typical for Phosphoglucomutase or Phosphomannomutase ; conversion of Hexose phosphates.	Metabolism
hfkd2_3o17	Strong similarity NADH Oxidoreductase B22 subunit-	The new protein is the human orthologe of the bovine EC 1.6.5.3. chain CI-B22 and therefore part of the human respiratory chain.	Metabolism
hfkd2_46j20	Strong similarity to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	The new protein seems to be the human 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.	Metabolism
hte83_15c24	Strong similarity to 2-hydroxyacid dehydrogenases	New human 2-hydroxyacid dehydrogenase	Metabolism
htes3_17l17	Strong similarity to transketolases	Transketolase testis specific	Metabolism
htes3_27d1	Similarity to ubiquitin-specific proteases	Protease	Metabolism
htes3_2a17	Similar to thiol-proteases	Putative thiol-protease	Metabolism
htes3_35b5	Strong similarity to bovine vacuolar ATPase (EC 3.6.1.-) chain A	ATPase	Metabolism
htes3_35k16	Similarity to acyl-CoA synthetase	Acyl-CoA synthetase	Metabolism
htes3_35n12	Strong similarity to ADP/ATP carrier proteins	Involved in mitochondrial energy metabolism	Metabolism
htes3_35n9	Carboxylesterase, splice variant	Carboxylesterase	Metabolism
hute1_20b19	Similarity to sarcosine oxidases	Sarcosine oxidases	Metabolism
hute1_20m24	Strong similarity to S.cerevisiae Alg9p probable mannosyltransferase	Possible mannosyltransferase	Metabolism
hute1_23e13	Strong similarity to heat shock 27K proteins	Heat shock protein related new subtilase	Metabolism

Group transcription factors

CloneID DKFZp...	Homology	Function	Group
hfkd2_46k19	Strong similarity to pterin-4-alpha-carbinolamine dehydratase	Dcoh is a bifunctional protein, complexed with biopterin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the biopterin cofactor of phenylalanine hydroxylase	Transcription factor
hfkd2_47a4	Similarity to zinc fingers	New putative transcription factor with one C2H2 zinc fingers.	Transcription factor
htes3_2e12	Similarity to finger proteins	Transcription factor with three C2H2 zinc fingers. Additionally, a cytochrome C family heme-binding site signature is present in the protein	Transcripton factros
htes3_21j15	3 strong similarity to "NY-CO-33"	Transcription factor	Transcription factors
htes3_17n12	Nearly identical to mouse SOX-LZ	SOX-LZ, related to SRY and HMG-box-Proteins	Transcription factors
hutel_18i19	Similarity to transcription factor SF3	The SREBP-2 protein is cleaved to release soluble NH2-terminal that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis; a lim domain; shows similarity to the common sunflower transcripti	Transcription factor
hutel_1i2	Similarity to Dictostelium myosin heavy chain kinase	Zn-finger protein	Transcription factor

Group uterus associated

CloneID DKFZp...	Homology	Function	Group
hutel_17k7	Similarity to HPBRII-4 MRNA	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_18c12	Similarity to candidate tumor suppressor p33ING1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_18i4	Weak similarity to C.elegans D2085.2	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_19g19	Partial similarity to bovine elastin fragment	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_19j11	Strong similarity to KIAA0231, similarity to ras binding protein Sur8	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_22n2	Similar to F46F6.1	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_21d15	Unknown	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_22o2	Similarity to S.pombe SPBC3E7.03c	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated
hutel_23g11	Similarity to SPAC31G5.12c and Maf1p	No informative BLAST results; No predictive prosite, pfam or SCOP motive	Uterus associated

Prosite Key

NAME: N-glycosylation site.
CONSENSUS: N-{P}-[ST]-{P}.

NAME: Glycosaminoglycan attachment site.
CONSENSUS: S-G-x-G.

NAME: Tyrosine sulfation site.

NAME: cAMP- and cGMP-dependent protein kinase phosphorylation site.
CONSENSUS: [RK](2)-x-[ST].

NAME: Protein kinase C phosphorylation site.
CONSENSUS: [ST]-x-[RK].

NAME: Casein kinase II phosphorylation site.
CONSENSUS: [ST]-x(2)-[DE].

NAME: Tyrosine kinase phosphorylation site.
CONSENSUS: [RK]-x(2,3)-[DE]-x(2,3)-Y.

NAME: N-myristoylation site.
CONSENSUS: G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}.

NAME: Amidation site.
CONSENSUS: x-G-[RK]-[RK].

NAME: Aspartic acid and asparagine hydroxylation site.
CONSENSUS: C-x-[DN]-x(4)-[FY]-x-C-x-C.

NAME: Vitamin K-dependent carboxylation domain.
CONSENSUS: x(12)-E-x(3)-E-x-C-x(6)-[DEN]-x-[LIVMFY]-x(9)-[FYW].

NAME: Phosphopantetheine attachment site.
CONSENSUS: [DEQGSTALMKRH]-[LIVMFYSTAC]-[GNQ]-[LIVMFYAG]-[DNEKHS]-S-[LIVMST]-
CONSENSUS: {PCFY}-[STAGCPQLIVMF]-[LIVMATN]-[DENQGTAKRHLM]-[LIVMWSTA]-[LIVGSTACR]-
CONSENSUS: x(2)-[LIVMFA].

NAME: Acyl carrier protein phosphopantetheine domain profile.

NAME: Prokaryotic membrane lipoprotein lipid attachment site.
CONSENSUS: {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C.

NAME: Prokaryotic N-terminal methylation site.
CONSENSUS: [KRHEQSTAG]-G-[FYLIVM]-[ST]-[LT]-[LIVP]-E-[LIVMFWSTAG](14).

NAME: Prenyl group binding site (CAAX box).
CONSENSUS: C-{DENQ}-[LIVM]-x>.

NAME: Protein splicing signature.
CONSENSUS: [DNEG]-x-[LIVFA]-[LIVMY]-[LVAST]-H-N-[STC].

NAME: Endoplasmic reticulum targeting sequence.
CONSENSUS: [KRHQSA]-[DENQ]-E-L>.

NAME: Microbodies C-terminal targeting signal.
CONSENSUS: [STAGCN]-[RKH]-[LIVMAFY]>.

NAME: Gram-positive cocci surface proteins 'anchoring' hexapeptide.
CONSENSUS: L-P-x-T-G-[STGAVDE].

NAME: Bipartite nuclear targeting sequence.

NAME: Cell attachment sequence.
CONSENSUS: R-G-D.

NAME: ATP/GTP-binding site motif A (P-loop).
CONSENSUS: [AG]-x(4)-G-K-[ST].

NAME: Cyclic nucleotide-binding domain signature 1.
CONSENSUS: [LIVM]-[VIC]-x(2)-G-[DENQTA]-x-[GAC]-x(2)-[LIVMFY](4)-x(2)-G.

NAME: Cyclic nucleotide-binding domain signature 2.
CONSENSUS: [LIVMF]-G-E-x-[GAS]-[LIVM]-x(5,11)-R-[STAQ]-A-x-[LIVMA]-x-[STACV].

NAME: cAMP/cGMP binding motif.

NAME: EF-hand calcium-binding domain.
CONSENSUS: D-x-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-
CONSENSUS: [DE]-[LIVMFYW].

NAME: Actinin-type actin-binding domain signature 1.
CONSENSUS: [EQ]-x(2)-[ATV]-[FY]-x(2)-W-x-N.

NAME: Actinin-type actin-binding domain signature 2.
CONSENSUS: [LIVM]-x-[SGN]-[LIVM]-[DAGHE]-[SAG]-x-[DNEAG]-[LIVM]-x-[DEAG]-x(4)-
CONSENSUS: [LIVM]-x-[LM]-[SAG]-[LIVM]-[LIVMT]-W-x-[LIVM](2).

NAME: Anaphylatoxin domain signature.
CONSENSUS: [CSH]-C-x(2)-[GAP]-x(7,8)-[GASTDEQR]-C-[GASTDEQL]-x(3,9)-[GASTDEQN]-x(2)-
CONSENSUS: [CE]-x(6,7)-C-C.

NAME: Anaphylatoxin domain profile.

NAME: Apple domain.
CONSENSUS: C-x(3)-[LIVMFY]-x(5)-[LIVMFY]-x(3)-[DENQ]-[LIVMFY]-x(10)-C-x(3)-C-T-
CONSENSUS: x(4)-C-x-[LIVMFY]-F-x-[FY]-x(13,14)-C-x-[LIVMFY]-[RK]-x-[ST]-x(14,15)-
CONSENSUS: S-G-x-[ST]-[LIVMFY]-x(2)-C.

NAME: Band 4.1 family domain signature 1.
CONSENSUS: W-[LIV]-x(3)-[KRQ]-x-[LIVM]-x(2)-[QH]-x(0,2)-[LIVMF]-x(6,8)-[LIVMF]-
CONSENSUS: x(3,5)-F-[FY]-x(2)-[DENS].

NAME: Band 4.1 family domain signature 2.
CONSENSUS: [HYW]-x(9)-[DENQSTV]-[SA]-x(3)-[FY]-[LIVM]-x(2)-[ACV]-x(2)-[LM]-x(2)-
CONSENSUS: [FY]-G-x-[DENQST]-[LIVMFYS].

NAME: Band 4.1 family domain profile.

NAME: C1q domain signature.
CONSENSUS: F-x(5)-[ND]-x(4)-[FYWL]-x(6)-F-x(5)-G-x-Y-x-F-x-[FY].

NAME: C-terminal cystine knot signature.
CONSENSUS: C-C-x(13)-C-x(2)-[GN]-x(12)-C-x-C-x(2,4)-C.

NAME: C-terminal cystine knot profile.

NAME: CUB domain profile.

NAME: Death domain profile.

NAME: EGF-like domain signature 1.
CONSENSUS: C-x-C-x(5)-G-x(2)-C.

NAME: EGF-like domain signature 2.
CONSENSUS: C-x-C-x(2)-[GP]-[FYW]-x(4,8)-C.

NAME: Calcium-binding EGF-like domain pattern signature.
CONSENSUS: [DEQN]-x-[DEQN](2)-C-x(3,14)-C-x(3,7)-C-x-[DN]-x(4)-[FY]-x-C.

NAME: Laminin-type EGF-like (LE) domain signature.
CONSENSUS: C-x(1,2)-C-x(5)-G-x(2)-C-x(2)-C-x(3,4)-[FYW]-x(3,15)-C.

NAME: Coagulation factors 5/8 type C domain (FA58C) signature 1.
CONSENSUS: [GAS]-W-x(7,15)-[FYW]-[LIV]-x-[LIVFA]-[GSTDEN]-x(6)-[LIVF]-x(2)-[IV]-x-
CONSENSUS: [LIVT]-[QKM]-G.

NAME: Coagulation factors 5/8 type C domain (FA58C) signature 2.
CONSENSUS: P-x(8,10)-[LM]-R-x-[GE]-[LIVP]-x-G-C.

NAME: Forkhead-associated (FHA) domain profile.

NAME: Fibrinogen beta and gamma chains C-terminal domain signature.
CONSENSUS: W-W-[LIVMFYW]-x(2)-C-x(2)-[GSA]-x(2)-N-G.

NAME: Type I fibronectin domain.

CONSENSUS: C-x(6,8)-[LFY]-x(5)-[FYW]-x-[RK]-x(8,10)-C-x-C-x(6,9)-C.

NAME: Type II fibronectin collagen-binding domain.
CONSENSUS: C-x(2)-P-F-x-[FYWI]-x(7)-C-x(8,10)-W-C-x(4)-[DNSR]-[FYW]-x(3,5)-[FYW]-x-
CONSENSUS: [FYWI]-C.

NAME: Hemopexin domain signature.
CONSENSUS: [LIFAT]-x(3)-W-x(2,3)-[PE]-x(2)-[LIVMFY]-[DENQS]-[STA]-[AV]-[LIVMFY].

NAME: Kringle domain signature.
CONSENSUS: [FY]-C-R-N-P-[DNR].

NAME: Kringle domain profile.

NAME: LDL-receptor class A (LDLRA) domain signature.
CONSENSUS: C-[VILMA]-x(5)-C-[DNH]-x(3)-[DENQHT]-C-x(3,4)-[STADE]-[DEH]-[DE]-x(1,5)-
CONSENSUS: C.

NAME: LDL-receptor class A (LDLRA) domain profile.

NAME: C-type lectin domain signature.
CONSENSUS: C-[LIVMFYATG]-x(5,12)-[WL]-x-[DNSR]-x(2)-C-x(5,6)-[FYWLIVSTA]-[LIVMSTA]-
CONSENSUS: C.

NAME: C-type lectin domain profile.

NAME: Link domain signature.
CONSENSUS: C-x(15)-A-x(3,4)-G-x(3)-C-x(2)-G-x(8,9)-P-x(7)-C.

NAME: Osteonectin domain signature 1.
CONSENSUS: C-x-[DN]-x(2)-C-x(2)-G-[KRH]-x-C-x(6,7)-P-x-C-x-C-x(3,5)-C-P.

NAME: Osteonectin domain signature 2.
CONSENSUS: F-P-x-R-[IM]-x-D-W-L-x-[NQ].

NAME: Somatomedin B domain signature.
CONSENSUS: C-x-C-x(3)-C-x(5)-C-C-x-[DN]-[FY]-x(3)-C.

NAME: Thyroglobulin type-1 repeat signature.
CONSENSUS: [FYWHP]-x-P-x-C-x(3,4)-G-x-[FYW]-x(3)-Q-C-x(4,10)-C-[FYW]-C-V-x(3,4)-
CONSENSUS: [SG].

NAME: P-type 'Trefoil' domain signature.
CONSENSUS: R-x(2)-C-x-[FYPST]-x(3,4)-[ST]-x(3)-C-x(4)-C-C-[FYWH].

NAME: Cellulose-binding domain, bacterial type.
CONSENSUS: W-N-[STAGR]-[STDN]-[LIVM]-x(2)-[GST]-x-[GST]-x(2)-[LIVMFT]-[GA].

NAME: Cellulose-binding domain, fungal type.
CONSENSUS: C-G-G-x(4,7)-G-x(3)-C-x(5)-C-x(3,5)-[NHG]-x-[FYWM]-x(2)-Q-C.

NAME: Chitin recognition or binding domain signature.
CONSENSUS: C-x(4,5)-C-C-S-x(2)-G-x-C-G-x(4)-[FYW]-C.

NAME: Barwin domain signature 1.
CONSENSUS: C-G-[KR]-C-L-x-V-x-N.

NAME: Barwin domain signature 2.
CONSENSUS: V-[DN]-Y-[EQ]-F-V-[DN]-C.

NAME: BIR repeat.
CONSENSUS: [HKEPILVY]-x(2)-R-x(3,7)-[FYW]-x(11,14)-[STAN]-G-[LMF]-X-[FYHDA]-X(4)-
CONSENSUS: [DESL]-X(2,3)-C-X(2)-C-X(6)-[WA]-X(9)-H-X(4)-[PRSD]-X-C-X(2)-[LIVMA].

NAME: WAP-type 'four-disulfide core' domain signature.
CONSENSUS: C-x-{C}-[DN]-x(2)-C-x(5)-C-C.

NAME: Phorbol esters / diacylglycerol binding domain.
CONSENSUS: H-x-[LIVMFYW]-x(8,11)-C-x(2)-C-x(3)-[LIVMFC]-x(5,10)-C-x(2)-C-x(4)-[HD]-
CONSENSUS: x(2)-C-x(5,9)-C.

NAME: C2 domain signature.
CONSENSUS: [ACG]-x(2)-L-x(2,3)-D-x(1,2)-[NGSTLIF]-[GTMR]-x-[STAP]-D-[PA]-[FY].

NAME: C2-domain profile.

NAME: CAP-Gly domain signature.
CONSENSUS: G-x(8,10)-[FYW]-x-G-[LIVM]-x-[LIVMFY]-x(4)-G-K-[NH]-x-G-[STAR]-x(2)-G-
CONSENSUS: x(2)-[LY]-F.

NAME: Ly-6 / u-PAR domain signature.
CONSENSUS: [EQR]-C-[LIVMFYAH]-x-C-x(5,8)-C-x(3,8)-[EDNQSTV]-C-{C}-x(5)-C-
CONSENSUS: x(12,24)-C.

NAME: MAM domain signature.
CONSENSUS: G-x-[LIVMFY](2)-x(3)-[STA]-x(10,11)-[LV]-x(4)-[LIVMF]-x(6,7)-C-[LIVM]-x-
CONSENSUS: F-x-[LIVMFY]-x(3)-[GSC].

NAME: MAM domain profile.

NAME: PH domain profile.

NAME: Phosphotyrosine interaction domain (PID) profile.

NAME: Src homology 2 (SH2) domain profile.

NAME: Src homology 3 (SH3) domain profile.

NAME: VWFC domain signature.
CONSENSUS: C-x(2,3)-C-x-C-x(6,14)-C-x(3,4)-C-x(2,10)-C-x(9,16)-C-C-x(2,4)-C.

NAME: WW/rsp5/WWP domain signature.
CONSENSUS: W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.

NAME: WW/rsp5/WWP domain profile.

NAME: ZP domain signature.
CONSENSUS: [LIVMFYW]-x(7)-[STAPDNL]-x(3)-[LIVMFYW]-x-[LIVMFYW]-x-[LIVMFYW]-x(2)-C-
CONSENSUS: [LIVMFYW]-x-[ST]-[PSL]-x(2,4)-[DENS]-x-[STADNQLF]-x(6)-[LIVM](2)-x(3,4)-
CONSENSUS: C.

NAME: S-layer homology domain signature.
CONSENSUS: [LVFYT]-x-[DA]-x(2,5)-[DNGSATPHY]-[WYFPDA]-x(4)-[LIV]-x(2)-[GTALV]-
CONSENSUS: x(4,6)-[LIVFYC]-x(2)-G-x-[PGSTA]-x(2,3)-[MFYA]-x-[PGAV]-x(3,10)-[LIVMA]-
CONSENSUS: [STKR]-[RY]-x-[EQ]-x-[STALIVM].

NAME: 'Homeobox' domain signature.
CONSENSUS: [LIVMFYG]-[ASLVR]-x(2)-[LIVMSTACN]-x-[LIVM]-x(4)-[LIV]-[RKNQESTAIY]-
CONSENSUS: [LIVFSTNKH]-W-[FYVC]-x-[NDQTAH]-x(5)-[RKNAIMW].

NAME: 'Homeobox' domain profile.

NAME: 'Homeobox' antennapedia-type protein signature.
CONSENSUS: [LIVMFE]-[FY]-P-W-M-[KRQTA].

NAME: 'Homeobox' engrailed-type protein signature.
CONSENSUS: L-M-A-Q-G-L-Y-N.

NAME: 'Paired box' domain signature.
CONSENSUS: R-P-C-x(11)-C-V-S.

NAME: 'POU' domain signature 1.
CONSENSUS: [RKQ]-R-[LIM]-x-[LF]-G-[LIVMFY]-x-Q-x-[DNQ]-V-G.

NAME: 'POU' domain signature 2.
CONSENSUS: S-Q-[ST]-[TA]-I-[SC]-R-F-E-x-[LSQ]-x-[LI]-[ST].

NAME: Zinc finger, C2H2 type, domain.
CONSENSUS: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H.

NAME: Zinc finger, C3HC4 type (RING finger), signature.
CONSENSUS: C-x-H-x-[LIVMFY]-C-x(2)-C-[LIVMYA].

NAME: Nuclear hormones receptors DNA-binding region signature.
CONSENSUS: C-x(2)-C-x-[DE]-x(5)-[HN]-[FY]-x(4)-C-x(2)-C-x(2)-F-F-x-R.

NAME: GATA-type zinc finger domain.
CONSENSUS: C-x-[DN]-C-x(4,5)-[ST]-x(2)-W-[HR]-[RK]-x(3)-[GN]-x(3,4)-C-N-[AS]-C.

NAME: Poly(ADP-ribose) polymerase zinc finger domain signature.
CONSENSUS: C-[KR]-x-C-x(3)-I-x-K-x(3)-[RG]-x(16,18)-W-[FYH]-H-x(2)-C.

NAME: Poly(ADP-ribose) polymerase zinc finger domain profile.

NAME: Fungal Zn(2)-Cys(6) binuclear cluster domain signature.
CONSENSUS: [GASTPV]-C-x(2)-C-[RKHSTACW]-x(2)-[RKHQ]-x(2)-C-x(5,12)-C-x(2)-C-x(6,8)-
CONSENSUS: C.

NAME: Fungal Zn(2)-Cys(6) binuclear cluster domain profile.

NAME: Prokaryotic dksA/traR C4-type zinc finger.
CONSENSUS: C-[DES]-x-C-x(3)-I-x(3)-R-x(4)-P-x(4)-C-x(2)-C.

NAME: Copper-fist domain signature.
CONSENSUS: M-[LIVMF](3)-x(3)-K-[MY]-A-C-x(2)-C-I-[KR]-x-H-[KR]-x(3)-C-x-H-x(8)-
CONSENSUS: [KR]-x-[KR]-G-R-P.

NAME: Copper fist DNA binding domain profile.

NAME: Leucine zipper pattern.
CONSENSUS: L-x(6)-L-x(6)-L-x(6)-L.

NAME: bZIP transcription factors basic domain signature.
CONSENSUS: [KR]-x(1,3)-[RKSAQ]-N-x(2)-[SAQ](2)-x-[RKTAENQ]-x-R-x-[RK].

NAME: Myb DNA-binding domain repeat signature 1.
CONSENSUS: W-[ST]-x(2)-E-[DE]-x(2)-[LIV].

NAME: Myb DNA-binding domain repeat signature 2.
CONSENSUS: W-x(2)-[LI]-[SAG]-x(4,5)-R-x(8)-[YW]-x(3)-[LIVM].

NAME: Myc-type, 'helix-loop-helix' dimerization domain signature.
CONSENSUS: [DENSTAP]-K-[LIVMWAGSN]-{FYWCPHKR}-[LIVT]-[LIV]-x(2)-[STAV]-[LIVMSTAC]-x-
CONSENSUS: [VMFYH]-[LIVMTA]-{P}-{P}-[LIVMSR].

NAME: p53 tumor antigen signature.
CONSENSUS: M-C-N-S-S-C-M-G-G-M-N-R-R.

NAME: CBF-A/NF-YB subunit signature.
CONSENSUS: C-V-S-E-x-I-S-F-[LIVM]-T-[SG]-E-A-[SC]-[DE]-[KRQ]-C.

NAME: CBF-B/NF-YA subunit signature.
CONSENSUS: Y-V-N-A-K-Q-Y-x-R-I-L-K-R-R-x-A-R-A-K-L-E.

NAME: 'Cold-shock' DNA-binding domain signature.
CONSENSUS: [FY]-G-F-I-x(6,7)-[DER]-[LIVM]-F-x-H-x-[STKR]-x-[LIVMFY].

NAME: CTF/NF-I signature.
CONSENSUS: R-K-R-K-Y-F-K-K-H-E-K-R.

NAME: Ets-domain signature 1.
CONSENSUS: L-[FYW]-[QEDH]-F-[LI]-[LVQK]-x-[LI]-L.

NAME: Ets-domain signature 2.
CONSENSUS: [RKH]-x(2)-M-x-Y-[DENQ]-x-[LIVM]-[STAG]-R-[STAG]-[LI]-R-x-Y.

NAME: Ets-domain profile.

NAME: Fork head domain signature 1.
CONSENSUS: [KR]-P-[PTQ]-[FYLVQH]-S-[FY]-x(2)-[LIVM]-x(3,4)-[AC]-[LIM].

NAME: Fork head domain signature 2.
CONSENSUS: W-[QKR]-[NS]-S-[LIV]-R-H.

NAME: Fork head domain profile.

NAME: HSF-type DNA-binding domain signature.
CONSENSUS: L-x(3)-[FY]-K-H-x-N-x-[STAN]-S-F-[LIVM]-R-Q-L-[NH]-x-Y-x-[FYW]-[RKH]-K-
CONSENSUS: [LIVM].

NAME: Tryptophan pentad repeat (IRF family) signature.
CONSENSUS: W-x-[DNH]-x(5)-[LIVF]-x-[IV]-P-W-x-H-x(9,10)-[DE]-x(2)-[LIVF]-F-[KRQ]-x-

CONSENSUS: [WR]-A.

NAME: LIM domain signature.
CONSENSUS: C-x(2)-C-x(15,21)-[FYWH]-H-x(2)-[CH]-x(2)-C-x(2)-C-x(3)-[LIVMF].

NAME: LIM domain profile.

NAME: NF-kappa-B/Rel/dorsal domain signature.
CONSENSUS: F-R-Y-x-C-E-G.

NAME: MADS-box domain signature.
CONSENSUS: R-x-[RK]-x(5)-I-x-[DN]-x(3)-[KR]-x(2)-T-[FY]-x-[RK](3)-x(2)-[LIVM]-x-
CONSENSUS: K(2)-A-x-E-[LIVM]-[ST]-x-L-x(4)-[LIVM]-x-[LIVM](3)-x(6)-[LIVMF]-x(2)-
CONSENSUS: [FY].

NAME: MADS-box domain profile.

NAME: T-box domain signature 1.
CONSENSUS: L-W-x(2)-[FC]-x(3,4)-[NT]-E-M-[LIV](2)-T-x(2)-G-[RG]-[KRQ].

NAME: T-box domain signature 2.
CONSENSUS: [LIVMYW]-H-[PADH]-[DEN]-[GS]-x(3)-G-x(2)-W-M-x(3)-[IVA]-x-F.

NAME: TEA domain signature.
CONSENSUS: G-R-N-E-L-I-x(2)-Y-I-x(3)-[TC]-x(3)-R-T-[RK](2)-Q-[LIVM]-S-S-H-[LIVM]-
CONSENSUS: Q-V.

NAME: Transcription factor TFIIB repeat signature.
CONSENSUS: G-[KR]-x(3)-[STAGN]-x-[LIVMYA]-[GSTA](2)-[CSAV]-[LIVM]-[LIVMFY]-[LIVMA]-
CONSENSUS: [GSA]-[STAC].

NAME: Transcription factor TFIID repeat signature.
CONSENSUS: Y-x-P-x(2)-[IF]-x(2)-[LIVM](2)-x-[KRH]-x(3)-P-[RKQ]-x(3)-L-[LIVM]-F-x-
CONSENSUS: [STN]-G-[KR]-[LIVM]-x(3)-G-[TAGL]-[KR]-x(7)-[AGC]-x(7)-[LIVM].

NAME: TFIIS zinc ribbon domain signature.
CONSENSUS: C-x(2)-C-x(9)-[LIVMQSAR]-[QH]-[STQL]-[RA]-[SACR]-x-[DE]-[DET]-[PGSEA]-
CONSENSUS: x(6)-C-x(2,5)-C-x(3)-[FW].

NAME: TSC-22 / dip / bun family signature.
CONSENSUS: M-D-L-V-K-x-H-L-x(2)-A-V-R-E-E-V-E.

NAME: Prokaryotic transcription elongation factors signature 1.
CONSENSUS: [ST]-x(2)-[GS]-x(3)-[LI]-x(2)-E-L-x(2)-L-x(3,4)-R-x(2)-[IV]-x(3)-[LIV]-
CONSENSUS: x(6)-G-D-x(2)-E-N-[GSA]-x-Y.

NAME: Prokaryotic transcription elongation factors signature 2.
CONSENSUS: S-x(2)-S-P-[LIVM]-[AG]-x-[SAG]-[LIVM]-[LIVMY]-x(4)-[DG]-[DE].

NAME: DEAD-box subfamily ATP-dependent helicases signature.
CONSENSUS: [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN].

NAME: DEAH-box subfamily ATP-dependent helicases signature.
CONSENSUS: [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR].

NAME: Eukaryotic putative RNA-binding region RNP-1 signature.
CONSENSUS: [RK]-G-{EDRKHPCG}-[AGSCI]-[FY]-[LIVA]-x-[FYLM].

NAME: Fibrillarin signature.
CONSENSUS: [GST]-[LIVMAP]-V-Y-A-[IV]-E-[FY]-[SA]-x-R-x(2)-R-[DE].

NAME: MCM family signature.
CONSENSUS: G-[IVT]-[LVAC](2)-[IVT]-D-[DE]-[FL]-[DNST].

NAME: MCM family domain.

NAME: XPA protein signature 1.
CONSENSUS: C-x-[DE]-C-x(3)-[LIVMF]-x(1,2)-D-x(2)-L-x(3)-F-x(4)-C-x(2)-C.

NAME: XPA protein signature 2.
CONSENSUS: [LIVM](2)-T-[KR]-T-E-x-K-x-[DE]-Y-[LIVMF](2)-x-D-x-[DE].

NAME: XPG protein signature 1.
CONSENSUS: [VI]-[KRE]-P-x-[FYIL]-V-F-D-G-x(2)-[PIL]-x-[LVC]-K.

NAME: XPG protein signature 2.
CONSENSUS: [GS]-[LIVM]-[PER]-[FYS]-[LIVM]-x-A-P-x-E-A-[DE]-[PAS]-[QS]-[CLM].

NAME: Bacterial regulatory proteins, araC family signature.
CONSENSUS: [KRQ]-[LIVMA]-x(2)-[GSTALIV]-{FYWPGDN}-x(2)-[LIVMSA]-x(4,9)-[LIVMF]-
CONSENSUS: x(2)-[LIVMSTA]-[GSTACIL]-x(3)-[GANQRF]-[LIVMFY]-x(4,5)-[LFY]-x(3)-
CONSENSUS: [FYIVA]-{FYWHCM}-x(3)-[GSADENQKR]-x-[NSTAPKL]-[PARL].

NAME: Bacterial regulatory proteins, araC family DNA-binding domain profile.

NAME: Bacterial regulatory proteins, arsR family signature.
CONSENSUS: C-x(2)-D-[LIVM]-x(6)-[ST]-x(4)-S-[HYR]-[HQ].

NAME: Bacterial regulatory proteins, asnC family signature.
CONSENSUS: [GSTAP]-x(2)-[DNEA]-[LIVM]-[GSA]-x(2)-[LIVMFY]-[GN]-[LIVMST]-[ST]-x(6)-R-
CONSENSUS: [LVT]-x(2)-[LIVM]-x(3)-G.

NAME: Bacterial regulatory proteins, crp family signature.
CONSENSUS: [LIVM]-[STAG]-[RHNW]-x(2)-[LIM]-[GA]-x-[LIVMFYA]-[LIVSC]-[GA]-x-[STACN]-
CONSENSUS: x(2)-[MST]-x-[GSTN]-R-x-[LIVMF]-x(2)-[LIVMF].

NAME: Bacterial regulatory proteins, deoR family signature.
CONSENSUS: R-x(3)-[LIVM]-x(3)-[LIVM]-x(16,17)-[STA]-x(2)-T-[LIVMA]-[RH]-[KRNA]-D-
CONSENSUS: [LIVMF].

NAME: Bacterial regulatory proteins, gntR family signature.
CONSENSUS: [LIVAPKR]-[PILV]-x-[EQTIVMR]-x(2)-[LIVM]-x(3)-[LIVMFYK]-x-[LIVFT]-
CONSENSUS: [DNGSTK]-[RGTLV]-x-[STAIVP]-[LIVA]-x(2)-[STAGV]-[LIVMFYH]-x(2)-[LMA].

NAME: Bacterial regulatory proteins, iclR family signature.
CONSENSUS: [GA]-x(3)-[DS]-x(2)-E-x(6)-[CSA]-[LIVM]-[GSA]-x(2)-[LIVM]-[FYH]-[DN].

NAME: Bacterial regulatory proteins, lacI family signature.
CONSENSUS: [LIVM]-x-[DE]-[LIVM]-A-x(2)-[STAGV]-x-V-[GSTP]-x(2)-[STAG]-[LIVMA]-x(2)-
CONSENSUS: [LIVMFYAN]-[LIVMC].

NAME: Bacterial regulatory proteins, luxR family signature.
CONSENSUS: [GDC]-x(2)-[NSTAVY]-x(2)-[IV]-[GSTA]-x(2)-[LIVMFYWCT]-x-[LIVMFYWCR]-x(3)-
CONSENSUS: [NST]-[LIVM]-x(5)-[NRHSA]-[LIVMSTA]-x(2)-[KR].

NAME: Bacterial regulatory proteins, lysR family signature.
CONSENSUS: [NQKRHSTAG]-[LIVMFYTA]-x(2)-[STAGLV]-[STAG]-x(4)-[LIVMYCTQR]-[PSTANLVER]-
CONSENSUS: x-[PSTAGQV]-[PSTAGNVMF]-[LIVMFA]-[STAGH]-x(2)-[LIVMF]-x(2)-[LIVMFW]-
CONSENSUS: [RKEAV]-x(2)-[LIVMFYNTAE]-x(3)-[LIMVT].

NAME: Bacterial regulatory proteins, marR family signature.
CONSENSUS: [STNA]-[LIA]-x-[RNGS]-x(4)-[LM]-[EIV]-x(2)-[GES]-[LFYW]-[LIVC]-x(7)-
CONSENSUS: [DN]-[RKQG]-[RK]-x(6)-T-x(2)-[GA].

NAME: Bacterial regulatory proteins, merR family signature.
CONSENSUS: [GSA]-x-[LIVMFA]-[ASM]-x(2)-[STACLIV]-[GSDENQR]-[LIVC]-[STANHK]-x(3)-
CONSENSUS: [LIVM]-[RHF]-x-[YW]-[DEQ]-x(2,3)-[GHDNQ]-[LIVMF](2).

NAME: Bacterial regulatory proteins, tetR family signature.
CONSENSUS: G-[LIVMFYS]-x(2,3)-[TS]-[LIVMT]-x(2)-[LIVM]-x(5)-[LIVQS]-[STAGENQH]-x-
CONSENSUS: [GPAR]-x-[LIVMF]-[FYST]-x-[HFY]-[FV]-x-[DNST]-K-x(2)-[LIVM].

NAME: Transcriptional antiterminators bglG family signature.
CONSENSUS: [ST]-x-H-x(2)-[FA](2)-[LIVM]-[EQK]-R-x(2)-[QNK].

NAME: Sigma-54 factors family signature 1.
CONSENSUS: P-[LIVM]-x-[LIVM]-x(2)-[LIVM]-A-x(2)-[LIVMF]-x(2)-[HS]-x-S-T-[LIVM]-S-R.

NAME: Sigma-54 factors family signature 2.
CONSENSUS: R-R-T-[IV]-[AT]-K-Y-R.

NAME: Sigma-54 factors family profile.

NAME: Sigma-70 factors family signature 1.
CONSENSUS: [DE]-[LIVMF](2)-[HEQS]-x-G-x-[LIVMFA]-G-L-[LIVMFYE]-x-[GSAM]-[LIVMAP].

NAME: Sigma-70 factors family signature 2.
CONSENSUS: [STN]-x(2)-[DEQ]-[LIVM]-[GAS]-x(4)-[LIVMF]-[PSTG]-x(3)-[LIVMA]-x-[NQR]-

CONSENSUS: [LIVMA]-[EQH]-x(3)-[LIVMFW]-x(2)-[LIVM].

NAME: Sigma-70 factors ECF subfamily signature.
CONSENSUS: [STAIV]-[PQDEL]-[DE]-[LIV]-[LIVTA]-Q-x-[STAV]-[LIVMFYC]-[LIVMAK]-x-
CONSENSUS: [GSTAIV]-[LIMFYWQ]-x(12,14)-[STAP]-[FYW]-[LIF]-x(2)-[IV].

NAME: Sigma-54 interaction domain ATP-binding region A signature.
CONSENSUS: [LIVMFY](3)-x-G-[DEQ]-[STE]-G-[STAV]-G-K-x(2)-[LIVMFY].

NAME: Sigma-54 interaction domain ATP-binding region B signature.
CONSENSUS: [GS]-x-[LIVMF]-x(2)-A-[DNEQASH]-[GNEK]-G-[STIM]-[LIVMFY](3)-[DE]-[EK]-
CONSENSUS: [LIVM].

NAME: Sigma-54 interaction domain C-terminal part signature.
CONSENSUS: [FYW]-P-[GS]-N-[LIVM]-R-[EQ]-L-x-[NHAT].

NAME: Sigma-54 interaction domain profile.

NAME: Single-strand binding protein family signature 1.
CONSENSUS: [LIVMF]-[NST]-[KRT]-[LIVM]-x-[LIVMF](2)-G-[NHRK]-[LIVM]-[GST]-x-[DET].

NAME: Single-strand binding protein family signature 2.
CONSENSUS: T-x-W-[HY]-[RNS]-[LIVM]-x-[LIVMF]-[FY]-[NGKR].

NAME: Bacterial histone-like DNA-binding proteins signature.
CONSENSUS: [GSK]-F-x(2)-[LIVMF]-x(4)-[RKEQA]-x(2)-[RST]-x-[GA]-x-[KN]-P-x-T.

NAME: Dps protein family signature 1.
CONSENSUS: H-[FW]-x-[LIVM]-x-G-x(5)-[LV]-H-x(3)-[DE].

NAME: Dps protein family signature 2.
CONSENSUS: [LIVMFY]-[DH]-x-[LIVM]-[GA]-E-R-x(3)-[LIF]-[GDN]-x(2)-[PA].

NAME: DNA repair protein radC family signature.
CONSENSUS: H-N-H-P-S-G.

NAME: recA signature.
CONSENSUS: A-L-[KR]-[IF]-[FY]-[STA]-[STAD]-[LIVMQ]-R.

NAME: RecF protein signature 1.
CONSENSUS: P-[ED]-x(3)-[LIVM](2)-x-G-[GSAD]-P-x(2)-R-R-x-[FY]-[LIVM]-D.

NAME: RecF protein signature 2.
CONSENSUS: [LIVMFY](2)-x-D-x(2,3)-[SA]-[EH]-L-D-x(2)-[KRH]-x(3)-L.

NAME: RecR protein signature.
CONSENSUS: C-x(2)-C-x(3)-[ST]-x(4)-C-x-I-C-x(4)-R.

NAME: Histone H2A signature.
CONSENSUS: [AC]-G-L-x-F-P-V.

NAME: Histone H2B signature.
CONSENSUS: [KR]-E-[LIVM]-[EQ]-T-x(2)-[KR]-x-[LIVM](2)-x-[PAG]-[DE]-L-x-[KR]-H-A-
CONSENSUS: [LIVM]-[STA]-E-G.

NAME: Histone H3 signature 1.
CONSENSUS: K-A-P-R-K-Q-L.

NAME: Histone H3 signature 2.
CONSENSUS: P-F-x-[RA]-L-[VA]-[KRQ]-[DEG]-[IV].

NAME: Histone H4 signature.
CONSENSUS: G-A-K-R-H.

NAME: HMG1/2 signature.
CONSENSUS: [FI]-S-[KR]-K-C-S-[EK]-R-W-K-T-M.

NAME: HMG-I and HMG-Y DNA-binding domain (A+T-hook).
CONSENSUS: [AT]-x(1,2)-[RK](2)-[GP]-R-G-R-P-[RK]-x.

NAME: HMG14 and HMG17 signature.
CONSENSUS: R-R-S-A-R-L-S-A-[RK]-P.

NAME: Bromodomain signature.

CONSENSUS: [STANVF]-x(2)-F-x(4)-[DNS]-x(5,7)-[DENQTF]-Y-[HFY]-x(2)-[LIVMFY]-x(3)-
CONSENSUS: [LIVM]-x(4)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(2)-N-[SACF]-x(2)-[FY].

NAME: Bromodomain profile.

NAME: Chromo domain signature.
CONSENSUS: [FYL]-x-[LIVMC]-[KR]-W-x-[GDNR]-[FYWLE]-x(5,6)-[ST]-W-[ES]-[PSTDN]-x(3)-
CONSENSUS: [LIVMC].

NAME: Chromo and chromo shadow domain profile.

NAME: Regulator of chromosome condensation (RCC1) signature 1.
CONSENSUS: G-x-N-D-x(2)-[AV]-L-G-R-x-T.

NAME: Regulator of chromosome condensation (RCC1) signature 2.
CONSENSUS: [LIVMFA]-[STAGC](2)-G-x(2)-H-[STAGLI]-[LIVMFA]-x-[LIVM].

NAME: Protamine P1 signature.
CONSENSUS: [AV]-R-[NFY]-R-x(2,3)-[ST]-x-S-x-S.

NAME: Nuclear transition protein 1 signature.
CONSENSUS: S-K-R-K-Y-R-K.

NAME: Nuclear transition protein 2 signature 1.
CONSENSUS: H-x(3)-H-S-[NS]-S-x-P-Q-S.

NAME: Nuclear transition protein 2 signature 2.
CONSENSUS: K-x-R-K-x(2)-E-G-K-x(2)-K-[KR]-K.

NAME: Ribosomal protein L1 signature.
CONSENSUS: [IM]-x(2)-[LIVA]-x(2,3)-[LIVM]-G-x(2)-[LMS]-[GSNH]-[PTKR]-[KRAV]-G-x-
CONSENSUS: [LMF]-P-[DENSTK].

NAME: Ribosomal protein L2 signature.
CONSENSUS: P-x(2)-R-G-[STAIV](2)-x-N-[APK]-x-[DE].

NAME: Ribosomal protein L3 signature.
CONSENSUS: [FL]-x(6)-[DN]-x(2)-[AGS]-x-[ST]-x-G-[KRH]-G-x(2)-G-x(3)-R.

NAME: Ribosomal protein L5 signature.
CONSENSUS: [LIVM]-x(2)-[LIVM]-[STAC]-[GE]-[QV]-x(2)-[LIVMA]-x-[STC]-x-[STAG]-[KR]-
CONSENSUS: x-[STA].

NAME: Ribosomal protein L6 signature 1.
CONSENSUS: [PS]-[DENS]-x-Y-K-[GA]-K-G-[LIVM].

NAME: Ribosomal protein L6 signature 2.
CONSENSUS: Q-x(3)-[LIVM]-x(2)-[KR]-x(2)-R-x-F-x-D-G-[LIVM]-Y-[LIVM]-x(2)-[KR].

NAME: Ribosomal protein L9 signature.
CONSENSUS: G-x(2)-[GN]-x(4)-V-x(2)-G-[FY]-x(2)-N-[FY]-L-x(5)-[GA]-x(3)-[STN].

NAME: Ribosomal protein L10 signature.
CONSENSUS: [DEH]-x(2)-[GS]-[LIVMF]-[STN]-[VA]-x-[DEQK]-[LIVMA]-x(2)-[LIM]-R.

NAME: Ribosomal protein L11 signature.
CONSENSUS: [RKN]-x-[LIVM]-x-G-[ST]-x(2)-[SNQ]-[LIVM]-G-x(2)-[LIVM]-x(0,1)-[DENG].

NAME: Ribosomal protein L13 signature.
CONSENSUS: [LIVM]-[KRV]-[GK]-M-[LIV]-[PS]-x(4,5)-[GS]-[NQEKRA]-x(5)-[LIVM]-x-[AIV]-
CONSENSUS: [LFY]-x-[GDN].

NAME: Ribosomal protein L14 signature.
CONSENSUS: [GA]-[LIV](3)-x(9,10)-[DNS]-G-x(4)-[FY]-x(2)-[NT]-x(2)-V-[LIV].

NAME: Ribosomal protein L15 signature.
CONSENSUS: K-[LIVM](2)-[GAL]-x-[GT]-x-[LIVMA]-x(2,5)-[LIVM]-x-[LIVMF]-x(3,4)-
CONSENSUS: [LIVMFC]-[ST]-x(2)-A-x(3)-[LIVM]-x(3)-G.

NAME: Ribosomal protein L16 signature 1.
CONSENSUS: [KR]-R-x-[GSAC]-[KQVA]-[LIVM]-W-[LIVM]-[KR]-[LIVM]-[LFY]-[AP].

NAME: Ribosomal protein L16 signature 2.
CONSENSUS: R-M-G-x-[GR]-K-G-x(4)-[FWKR].

NAME: Ribosomal protein L17 signature.
CONSENSUS: I-x-[ST]-[GT]-x(2)-[KR]-x-K-x(6)-[DE]-x-[LIMV]-[LIVMT]-T-x-[STAG]-[KR].

NAME: Ribosomal protein L19 signature.
CONSENSUS: [RT]-[KRSVY]-[GSA]-x-V-[RS]-[KR]-[SA]-K-L-Y-Y-L-R.

NAME: Ribosomal protein L20 signature.
CONSENSUS: K-x(3)-[KRC]-x-[LIVM]-W-[IV]-[STNALV]-R-[LIVM]-N-x(3)-[RKH].

NAME: Ribosomal protein L21 signature.
CONSENSUS: [IVT]-x(3)-[KR]-x(3)-[KRQ]-K-x(6)-G-[HF]-R-[RQ]-x(2)-T.

NAME: Ribosomal protein L22 signature.
CONSENSUS: [RKQN]-x(4)-[RH]-[GAS]-x-G-[KRQS]-x(9)-[HDN]-[LIVM]-x-[LIVMS]-x-[LIVM].

NAME: Ribosomal protein L23 signature.
CONSENSUS: [RK](2)-[AM]-[IVFYT]-[IV]-[RKT]-L-[STANQK]-x(7)-[LIVMFT].

NAME: Ribosomal protein L24 signature.
CONSENSUS: [GDEN]-D-x-V-x-[IV]-[LIVMA]-x-G-x(2)-[KA]-[GN]-x(2,3)-[GA]-x-[IV].

NAME: Ribosomal protein L27 signature.
CONSENSUS: G-x-[LIVM](2)-x-R-Q-R-G-x(5)-G.

NAME: Ribosomal protein L29 signature.
CONSENSUS: [KNQS]-[PSTL]-x(2)-[LIMFA]-[KRGSAN]-x-[LIVYSTA]-[KR]-[KRH]-[DESTANRL]-
CONSENSUS: [LIV]-A-[KRCQVT]-[LIVMA].

NAME: Ribosomal protein L30 signature.
CONSENSUS: [IVT]-[LIVM]-x(2)-[LF]-x-[LI]-x-[KRHQEG]-x(2)-[STNQH]-x-[IVT]-
CONSENSUS: x(10)-[LMS]-[LIV]-x(2)-[LIVA]-x(2)-[LMFY]-[IVT].

NAME: Ribosomal protein L31 signature.
CONSENSUS: H-P-F-[FY]-[TI]-x(9)-G-R-[AV]-x-[KR].

NAME: Ribosomal protein L33 signature.
CONSENSUS: Y-x-[ST]-x-[KR]-[NS]-x(4)-[PAT]-x(1,2)-[LIVM]-[EA]-x(2)-K-[FY]-[CSD].

NAME: Ribosomal protein L34 signature.
CONSENSUS: K-[RG]-T-[FYWL]-[EQS]-x(5)-[KRHS]-x(4,5)-G-F-x(2)-R.

NAME: Ribosomal protein L35 signature.
CONSENSUS: [LIVM]-K-[TV]-x(2)-[GSA]-[SAIL]-x-K-R-[LIVMFY]-[KRL].

NAME: Ribosomal protein L36 signature.
CONSENSUS: C-x(2)-C-x(2)-[LIVM]-x-R-x(3)-[LIVMN]-x-[LIVM]-x-C-x(3,4)-[KR]-H-x-Q-x-Q.

NAME: Ribosomal protein L1e signature.
CONSENSUS: N-x(3)-[KR]-x(2)-A-[LIVT]-x-S-A-[LIV]-x-A-[ST]-[SGA]-x(7)-[RK]-G-H.

NAME: Ribosomal protein L6e signature.
CONSENSUS: N-x(2)-P-L-R-R-x(4)-[FY]-V-I-A-T-S-x-K.

NAME: Ribosomal protein L7Ae signature.
CONSENSUS: [CA]-x(4)-[IV]-P-[FY]-x(2)-[LIVM]-x-[GSQ]-[KRQ]-x(2)-L-G.

NAME: Ribosomal protein L10e signature.
CONSENSUS: R-x-A-[FYW]-G-K-[PA]-x-G-x(2)-A-R-V.

NAME: Ribosomal protein L13e signature.
CONSENSUS: [KR]-Y-x(2)-K-[LIVM]-R-[STA]-G-[KR]-G-F-[ST]-L-x-E.

NAME: Ribosomal protein L15e signature.
CONSENSUS: [DE]-[KR]-A-R-x-L-G-[FY]-x-[SAP]-x(2)-G-[LIVMFY](4)-R-x-R-V-x-R-G.

NAME: Ribosomal protein L18e signature.
CONSENSUS: [KRE]-x-L-x(2)-[PS]-[KR]-x(2)-[RH]-[PSA]-x-[LIVM]-[NS]-[LIVM]-x-[RK]-
CONSENSUS: [LIVM].

NAME: Ribosomal protein L19e signature.
CONSENSUS: R-x-[KR]-x(5)-[KR]-x(3)-[KRH]-x(2)-G-x-G-x-R-x-G-x(3)-A-R-x(3)-[KQ]-
CONSENSUS: x(2)-W-x(7)-R-x(2)-L-x(3)-R.

NAME: Ribosomal protein L21e signature.
CONSENSUS: G-[DE]-x-V-x(10)-[GV]-x(2)-[FYH]-x(2)-[FY]-x-G-x-T-G.

NAME: Ribosomal protein L24e signature.
CONSENSUS: [FY]-x-[GS]-x(2)-[IV]-x-P-G-x-G-x(2)-[FYV]-x-[KRHE]-x-D.

NAME: Ribosomal protein L27e signature.
CONSENSUS: G-K-N-x-W-F-F-x-K-L-R-F>.

NAME: Ribosomal protein L30e signature 1.
CONSENSUS: [STA]-x(5)-G-x-[QKR]-x(2)-[LIVM]-[KQT]-x(2)-[KR]-x-G-x(2)-K-x-[LIVM](3).

NAME: Ribosomal protein L30e signature 2.
CONSENSUS: [DE]-L-G-[STA]-x(2)-G-[KR]-x(6)-[LIVM]-x-[LIVM]-x-[DEN]-x-G.

NAME: Ribosomal protein L31e signature.
CONSENSUS: V-[KR]-[LIVM]-x(3)-[LIVM]-N-x-[AK]-x-W-x-[KR]-G.

NAME: Ribosomal protein L32e signature.
CONSENSUS: F-x-R-x(4)-[KR]-x(2)-[KR]-[LIVM]-x(3)-W-R-[KR]-x(2)-G.

NAME: Ribosomal protein L34e signature.
CONSENSUS: Y-x-[ST]-x-S-[NY]-x(5)-[KR]-T-P-G.

NAME: Ribosomal protein L35Ae signature.
CONSENSUS: G-K-[LIVM]-x-R-x-H-G-x(2)-G-x-V-x-A-x-F-x(3)-[LI]-P.

NAME: Ribosomal protein L36e signature.
CONSENSUS: P-Y-E-[KR]-R-x-[LIVM]-[DE]-[LIVM](2)-[KR].

NAME: Ribosomal protein L37e signature.
CONSENSUS: G-T-x-[SA]-x-G-x-[KR]-x(3)-[ST]-x(0,1)-H-x(2)-C-x-R-C-G.

NAME: Ribosomal protein L39e signature.
CONSENSUS: [KRA]-T-x(3)-[LIVM]-[KRQF]-x-[NHS]-x(3)-R-[NHY]-W-R-R.

NAME: Ribosomal protein L44e signature.
CONSENSUS: K-x-[TV]-K-K-x(2)-L-[KR]-x(2)-C.

NAME: Ribosomal protein S2 signature 1.
CONSENSUS: [LIVMFA]-x(2)-[LIVMFYC](2)-x-[STAC]-[GSTANQEKR]-[STALV]-[HY]-[LIVMF]-G.

NAME: Ribosomal protein S2 signature 2.
CONSENSUS: P-x(2)-[LIVMF](2)-[LIVMS]-x-[GDN]-x(3)-[DENL]-x(3)-[LIVM]-x-E-x(4)-
CONSENSUS: [GNQKRH]-[LIVM]-[AP].

NAME: Ribosomal protein S3 signature.
CONSENSUS: [GSTA]-[KR]-x(6)-G-x-[LIVMT]-x(2)-[NQSCH]-x(1,3)-[LIVFCA]-x(3)-[LIV]-
CONSENSUS: [DENQ]-x(7)-[LMT]-x(2)-G-x(2)-G.

NAME: Ribosomal protein S4 signature.
CONSENSUS: [LIVM]-[DE]-x-R-L-x(3)-[LIVMC]-[VMFYHQ]-[KRT]-x(3)-[STAGCF]-x-[ST]-x(3)-
CONSENSUS: [SAI]-[KR]-x-[LIVMF](2).

NAME: Ribosomal protein S5 signature.
CONSENSUS: G-[KRQ]-x(3)-[FY]-x-[ACV]-x(2)-[LIVMA]-[LIVM]-[AG]-[DN]-x(2)-G-x-
CONSENSUS: [LIVM]-G-x-[SAG]-x(5,6)-[DEQ]-[LIVM]-x(2)-A-[LIVMF].

NAME: Ribosomal protein S6 signature.
CONSENSUS: G-x-[KRC]-[DENQRH]-L-[SA]-Y-x-I-[KRNSA].

NAME: Ribosomal protein S7 signature.
CONSENSUS: [DENSK]-x-[LIVMET]-x(3)-[LIVMFT](2)-x(6)-G-K-[KR]-x(5)-[LIVMF]-[LIVMFC]-
CONSENSUS: x(2)-[STA].

NAME: Ribosomal protein S8 signature.
CONSENSUS: [GE]-x(2)-[LIV](2)-[STY]-T-x(2)-G-[LIVM](2)-x(4)-[AG]-[KRHAYI].

NAME: Ribosomal protein S9 signature.
CONSENSUS: G-G-G-x(2)-[GSA]-Q-x(2)-[SA]-x(3)-[GSA]-x-[GSTAV]-[KR]-[GSAL]-[LIF].

NAME: Ribosomal protein S10 signature.
CONSENSUS: [AV]-x(3)-[GDNSR]-[LIVMSTA]-x(3)-G-P-[LIVM]-x-[LIVM]-P-T.

NAME: Ribosomal protein S11 signature.
CONSENSUS: [LIVMF]-x-[GSTAC]-[LIVMF]-x(2)-[GSTAL]-x(0,1)-[GSN]-[LIVMF]-x-[LIVM]-
CONSENSUS: x(4)-[DEN]-x-T-P-x-[PA]-[STCH]-[DN].

NAME: Ribosomal protein S12 signature.
CONSENSUS: [RK]-x-P-N-S-[AR]-x-R.

NAME: Ribosomal protein S13 signature.
CONSENSUS: [KRQS]-G-x-R-H-x(2)-[GSNH]-x(2)-[LIVMC]-R-G-Q.

NAME: Ribosomal protein S14 signature.
CONSENSUS: [RP]-x(0,1)-C-x(11,12)-[LIVMF]-x-[LIVMF]-[SC]-[RG]-x(3)-[RN].

NAME: Ribosomal protein S15 signature.
CONSENSUS: [LIVM]-x(2)-H-[LIVMFY]-x(5)-D-x(2)-[SAGN]-x(3)-[LF]-x(9)-[LIVM]-x(2)-
CONSENSUS: [FY].

NAME: Ribosomal protein S16 signature.
CONSENSUS: [LIVMT]-x-[LIVM]-[KR]-L-[STAK]-R-x-G-[AKR].

NAME: Ribosomal protein S17 signature.
CONSENSUS: G-D-x-[LIV]-x-[LIVA]-x-[QEK]-x-[RK]-P-[LIV]-S.

NAME: Ribosomal protein S18 signature.
CONSENSUS: [IV]-[DY]-Y-x(2)-[LIVMT]-x(2)-[LIVM]-x(2)-[FYT]-[LIVM]-[ST]-[DERP]-x-
CONSENSUS: [GY]-K-[LIVM]-x(3)-R-[LIVMAS].

NAME: Ribosomal protein S19 signature.
CONSENSUS: [STDNQ]-G-[KRQM]-x(6)-[LIVM]-x(4)-[LIVM]-[GSD]-x(2)-[LF]-[GAS]-[DE]-F-
CONSENSUS: x(2)-[ST].

NAME: Ribosomal protein S21 signature.
CONSENSUS: [DE]-x-A-[LY]-[KR]-R-F-K-[KR]-x(3)-[KR].

NAME: Ribosomal protein S3Ae signature.
CONSENSUS: [LIV]-x-[GH]-R-[IV]-x-E-x-[SC]-L-x-D-L.

NAME: Ribosomal protein S4e signature.
CONSENSUS: H-x-K-R-[LIVM]-[SAN]-x-P-x(2)-W-x-[LIVM]-x-[KR].

NAME: Ribosomal protein S6e signature.
CONSENSUS: [LIVM]-[STAMR]-G-G-x-D-x(2)-G-x-P-M.

NAME: Ribosomal protein S7e signature.
CONSENSUS: [KR]-L-x-R-E-L-E-K-K-F-[SAP]-x-[KR]-H.

NAME: Ribosomal protein S8e signature.
CONSENSUS: R-x(2)-T-G-[GA]-x(5)-[HR]-K-[KR]-x-K-x-E-[LM]-G.

NAME: Ribosomal protein S12e signature.
CONSENSUS: A-L-[KRQP]-x-V-L-x(2)-[SA]-x(3)-[DN]-G-L.

NAME: Ribosomal protein S17e signature.
CONSENSUS: A-x-I-x-[ST]-K-x-L-R-N-[KR]-I-A-G-[FY]-x-T-H.

NAME: Ribosomal protein S19e signature.
CONSENSUS: P-x(6)-[SAN]-x(2)-[LIVMA]-x-R-x-[ALIV]-[LV]-Q-x-L-[EQ].

NAME: Ribosomal protein S21e signature.
CONSENSUS: L-Y-V-P-R-K-C-S-[SA].

NAME: Ribosomal protein S24e signature.
CONSENSUS: [FA]-G-x(2)-[KR]-[STA]-x-G-[FY]-[GA]-x-[LIVM]-Y-[DN]-[SN].

NAME: Ribosomal protein S26e signature.
CONSENSUS: [YH]-C-V-S-C-A-I-H.

NAME: Ribosomal protein S27e signature.
CONSENSUS: [QK]-C-x(2)-C-x(6)-F-[GS]-x-[PSA]-x(5)-C-x(2)-C-[GS]-x(2)-L-x(2)-P-x-G.

NAME: Ribosomal protein S28e signature.
CONSENSUS: E-[ST]-E-R-E-A-R-x-L.

NAME: DNA mismatch repair proteins mutL / hexB / PMS1 signature.

CONSENSUS: G-F-R-G-E-A-L.

NAME: DNA mismatch repair proteins mutS family signature.
CONSENSUS: [ST]-[LIVM]-x-[LIVM]-x-D-E-[LIVMY]-[GC]-[RKH]-G-[GST]-x(4)-G.

NAME: mutT domain signature.
CONSENSUS: G-x(5)-E-x(4)-[STAGC]-[LIVMAC]-x-R-E-[LIVMFT]-x-E-E.

NAME: DnaA protein signature.
CONSENSUS: I-[GA]-x(2)-[LIVMF]-[SGDNK]-x(0,1)-[KR]-x-H-[STP]-[STV]-[LIVM](2)-x-
CONSENSUS: [SA]-x(2)-[KRE]-[LIVM].

NAME: Small, acid-soluble spore proteins, alpha/beta type, signature 1.
CONSENSUS: K-x-E-[LIV]-A-x-[DE]-[LIVMF]-G-[LIVMF].

NAME: Small, acid-soluble spore proteins, alpha/beta type, signature 2.
CONSENSUS: [KR]-[SAQ]-x-G-x-V-G-G-x-[LIVM]-x-[KR](2)-[LIVM](2).

NAME: Zinc-containing alcohol dehydrogenases signature.
CONSENSUS: G-H-E-x(2)-G-x(5)-[GA]-x(2)-[IVSAC].

NAME: Quinone oxidoreductase / zeta-crystallin signature.
CONSENSUS: [GSD]-[DEQH]-x(2)-L-x(3)-[SA](2)-G-G-x-G-x(4)-Q-x(2)-[KR].

NAME: Iron-containing alcohol dehydrogenases signature 1.
CONSENSUS: [STALIV]-[LIVF]-x-[DE]-x(6,7)-P-x(4)-[ALIV]-x-[GST]-x(2)-D-[TAIVM]-
CONSENSUS: [LIVMF]-x(4)-E.

NAME: Iron-containing alcohol dehydrogenases signature 2.
CONSENSUS: [GSW]-x-[LIVTSACD]-[GH]-x(2)-[GSAE]-[GSHYQ]-x-[LIVTP]-[GAST]-[GAS]-x(3)-
CONSENSUS: [LIVMT]-x-[HNS]-[GA]-x-[GTAC].

NAME: Short-chain dehydrogenases/reductases family signature.
CONSENSUS: [LIVSPADNK]-x(12)-Y-[PSTAGNCV]-[STAGNQCIVM]-[STAGC]-K-{PC}-[SAGFR]-
CONSENSUS: [LIVMSTAGD]-x(2)-[LIVMFYW]-x(3)-[LIVMFYWGAPTHQ]-[GSACQRHM].

NAME: Aldo/keto reductase family signature 1.
CONSENSUS: G-[FY]-R-[HSAL]-[LIVMF]-D-[STAGC]-[AS]-x(5)-E-x(2)-[LIVM]-G.

NAME: Aldo/keto reductase family signature 2.
CONSENSUS: [LIVMFY]-x(9)-[KREQ]-x-[LIVM]-G-[LIVM]-[SC]-N-[FY].

NAME: Aldo/keto reductase family putative active site signature.
CONSENSUS: [LIVM]-[PAIV]-[KR]-[ST]-x(4)-R-x(2)-[GSTAEQK]-[NSL]-x(2)-[LIVMFA].

NAME: Homoserine dehydrogenase signature.
CONSENSUS: A-x(3)-G-[LIVMFY]-[STAG]-x(2,3)-[DNS]-P-x(2)-D-[LIVM]-x-G-x-D-x(3)-K.

NAME: NAD-dependent glycerol-3-phosphate dehydrogenase signature.
CONSENSUS: G-[AT]-[LIVM]-K-[DN]-[LIVM](2)-A-x-[GA]-x-G-[LIVMF]-x-[DE]-G-[LIVM]-x-
CONSENSUS: [LIVMFYW]-G-x-N.

NAME: FAD-dependent glycerol-3-phosphate dehydrogenase signature 1.
CONSENSUS: [IV]-G-G-G-x(2)-G-[STACV]-G-x-A-x-D-x(3)-R-G.

NAME: FAD-dependent glycerol-3-phosphate dehydrogenase signature 2.
CONSENSUS: G-G-K-x(2)-[GSTE]-Y-R-x(2)-A.

NAME: Mannitol dehydrogenases signature.
CONSENSUS: [LIVMY]-x-[FS]-x(2)-[STAGCV]-x-V-D-R-[IV]-x-[PS].

NAME: Histidinol dehydrogenase signature.
CONSENSUS: I-D-x(2)-A-G-P-[ST]-E-[LIVS]-[LIVMA](3)-[AC]-x(3)-A-x(4)-[LIVM]-[AV]-
CONSENSUS: [SACL]-[DE]-[LIVMFC]-[LIVM]-[SA]-x(2)-E-H.

NAME: L-lactate dehydrogenase active site.
CONSENSUS: [LIVMA]-G-[EQ]-H-G-[DN]-[ST].

NAME: D-isomer specific 2-hydroxyacid dehydrogenases NAD-binding signature.
CONSENSUS: [LIVMA]-[AG]-[IVT]-[LIVMFY]-[AG]-x-G-[NHKRQGSAC]-[LIV]-G-x(13,14)-
CONSENSUS: [LIVfMT]-x(2)-[FYwCTH]-[DNSTK].

NAME: D-isomer specific 2-hydroxyacid dehydrogenases signature 2.
CONSENSUS: [LIVMFYWA]-[LIVFYWC]-x(2)-[SAC]-[DNQHR]-[IVFA]-[LIVF]-x-[LIVF]-[HNI]-x-

CONSENSUS: P-x(4)-[STN]-x(2)-[LIVMF]-x-[GSDN].

NAME: D-isomer specific 2-hydroxyacid dehydrogenases signature 3.
CONSENSUS: [LMFATC]-[KPQ]-x-[GSTDN]-x-[LIVMFYWR]-[LIVMFYW](2)-N-x-[STAGC]-R-[GP]-x-
CONSENSUS: [LIVH]-[LIVMC]-[DNV].

NAME: 3-hydroxyisobutyrate dehydrogenase signature.
CONSENSUS: [LIVMFY](2)-G-L-G-x-[MQ]-G-x-[PGS]-[MA]-[SA].

NAME: Hydroxymethylglutaryl-coenzyme A reductases signature 1.
CONSENSUS: [RKH]-x(6)-D-x-M-G-x-N-x-[LIVMA].

NAME: Hydroxymethylglutaryl-coenzyme A reductases signature 2.
CONSENSUS: [LIVM]-G-x-[LIVM]-G-G-[AG]-T.

NAME: Hydroxymethylglutaryl-coenzyme A reductases signature 3.
CONSENSUS: A-[LIVM]-x-[STAN]-x(2)-[LI]-x-[KRNQ]-[GSA]-H-[LM]-x-[FYLH].

NAME: Hydroxymethylglutaryl-coenzyme A reductases profile.

NAME: 3-hydroxyacyl-CoA dehydrogenase signature.
CONSENSUS: [DNE]-x(2)-[GA]-F-[LIVMFY]-x-[NT]-R-x(3)-[PA]-[LIVMFY](2)-x(5)-
CONSENSUS: [LIVMFYCT]-[LIVMFY]-x(2)-[GV].

NAME: Malate dehydrogenase active site signature.
CONSENSUS: [LIVM]-T-[TRKMN]-L-D-x(2)-R-[STA]-x(3)-[LIVMFY].

NAME: Malic enzymes signature.
CONSENSUS: F-x-[DV]-D-x(2)-G-T-[GSA]-x-[IV]-x-[LIVMA]-[GAST](2)-[LIVMF](2).

NAME: Isocitrate and isopropylmalate dehydrogenases signature.
CONSENSUS: [NS]-[LIMYT]-[FYDN]-G-[DNT]-[IMVY]-x-[STGDN]-[DN]-x(2)-[SGAP]-x(3,4)-G-
CONSENSUS: [STG]-[LIVMPA]-G-[LIVMF].

NAME: 6-phosphogluconate dehydrogenase signature.
CONSENSUS: [LIVM]-x-D-x(2)-[GA]-[NQS]-K-G-T-G-x-W.

NAME: Glucose-6-phosphate dehydrogenase active site.
CONSENSUS: D-H-Y-L-G-K-[EQK].

NAME: IMP dehydrogenase / GMP reductase signature.
CONSENSUS: [LIVM]-[RK]-[LIVM]-G-[LIVM]-G-x-G-S-[LIVM]-C-x-T.

NAME: Bacterial quinoprotein dehydrogenases signature 1.
CONSENSUS: [DEN]-W-x(3)-G-[RK]-x(6)-[FYW]-S-x(4)-[LIVM]-N-x(2)-N-V-x(2)-L-[RK].

NAME: Bacterial quinoprotein dehydrogenases signature 2.
CONSENSUS: W-x(4)-Y-D-x(3)-[DN]-[LIVMFY](4)-x(2)-G-x(2)-[STA]-P.

NAME: FMN-dependent alpha-hydroxy acid dehydrogenases active site.
CONSENSUS: S-N-H-G-[AG]-R-Q.

NAME: GMC oxidoreductases signature 1.
CONSENSUS: [GA]-[RKN]-x-[LIV]-G(2)-[GST](2)-x-[LIVM]-N-x(3)-[FYWA]-x(2)-[PAG]-x(5)-
CONSENSUS: [DNESH].

NAME: GMC oxidoreductases signature 2.
CONSENSUS: [GS]-[PSTA]-x(2)-[ST]-P-x-[LIVM](2)-x(2)-S-G-[LIVM]-G.

NAME: Eukaryotic molybdopterin oxidoreductases signature.
CONSENSUS: [GA]-x(3)-[KRNQHT]-x(11,14)-[LIVMFYWS]-x(8)-[LIVMF]-x-C-x(2)-[DEN]-R-
CONSENSUS: x(2)-[DE].

NAME: Prokaryotic molybdopterin oxidoreductases signature 1.
CONSENSUS: [STAN]-x-[CH]-x(2,3)-C-[STAG]-[GSTVMF]-x-C-x-[LIVMFYW]-x-[LIVMA]-x(3,4)-
CONSENSUS: [DENQKHT].

NAME: Prokaryotic molybdopterin oxidoreductases signature 2.
CONSENSUS: [STA]-x-[STAC](2)-x(2)-[STA]-D-[LIVMY](2)-L-P-x-[STAC](2)-x(2)-E.

NAME: Prokaryotic molybdopterin oxidoreductases signature 3.
CONSENSUS: A-x(3)-[GDT]-I-x-[DNQTK]-x-[DEA]-x-[LIVM]-x-[LIVMC]-x-[NS]-x(2)-[GS]-
CONSENSUS: x(5)-A-x-[LIVM]-[ST].

NAME: Aldehyde dehydrogenases glutamic acid active site.
CONSENSUS: [LIVMFGA]-E-[LIMSTAC]-[GS]-G-[KNLM]-[SADN]-[TAPFV].

NAME: Aldehyde dehydrogenases cysteine active site.
CONSENSUS: [FYLVA]-x(3)-G-[QE]-x-C-[LIVMGSTANC]-[AGCN]-x-[GSTADNEKR].

NAME: Aspartate-semialdehyde dehydrogenase signature.
CONSENSUS: [LIVM]-[SADN]-x(2)-C-x-R-[LIVM]-x(4)-[GSC]-H-[STA].

NAME: Glyceraldehyde 3-phosphate dehydrogenase active site.
CONSENSUS: [ASV]-S-C-[NT]-T-x(2)-[LIM].

NAME: N-acetyl-gamma-glutamyl-phosphate reductase active site.
CONSENSUS: [LIVM]-[GSA]-x-P-G-C-[FY]-[AVP]-T-[GA]-x(3)-[GTAC]-[LIVM]-x-P.

NAME: Gamma-glutamyl phosphate reductase signature.
CONSENSUS: V-x(5)-A-[LIV]-x-H-I-x(2)-[HY]-[GS]-[ST]-x-H-[ST]-[DE]-x-I.

NAME: Dihydrodipicolinate reductase signature.
CONSENSUS: E-[IV]-x-E-x-H-x(3)-K-x-D-x-P-S-G-T-A.

NAME: Dihydroorotate dehydrogenase signature 1.
CONSENSUS: [GS]-x(4)-[GK]-[STA]-[IVSTA]-[GT]-x(3)-[NQR]-x-G-[NH]-x(2)-P-[RT].

NAME: Dihydroorotate dehydrogenase signature 2.
CONSENSUS: [LIV](2)-[GSA]-x-G-G-[IV]-x-[STGN]-x(3)-[ACV]-x(6)-G-A.

NAME: Coproporphyrinogen III oxidase signature.
CONSENSUS: K-x-W-C-x(2)-[FYH](3)-[LIVM]-x-H-R-x-E-x-R-G-[LIVM]-G-G-[LIVM]-F-F-D.

NAME: Fumarate reductase / succinate dehydrogenase FAD-binding site.
CONSENSUS: R-[ST]-H-[ST]-x(2)-A-x-G-G.

NAME: Acyl-CoA dehydrogenases signature 1.
CONSENSUS: [GAC]-[LIVM]-[ST]-E-x(2)-[GSAN]-G-[ST]-D-x(2)-[GSA].

NAME: Acyl-CoA dehydrogenases signature 2.
CONSENSUS: [QDE]-x(2)-G-[GS]-x-G-[LIVMFY]-x(2)-[DEN]-x(4)-[KR]-x(3)-[DEN].

NAME: Alanine dehydrogenase & pyridine nucleotide transhydrogenase signature 1.
CONSENSUS: G-[LIVM]-P-x-E-x(3)-N-E-x(1,3)-R-V-A-x-[ST]-P-x-[GST]-V-x(2)-L-x-[KRH]-
CONSENSUS: x-G.

NAME: Alanine dehydrogenase & pyridine nucleotide transhydrogenase signature 2.
CONSENSUS: [LIVM](2)-G-[GA]-G-x-A-G-x(2)-[SA]-x(3)-[GA]-x-[SG]-[LIVM]-G-A-x-V-
CONSENSUS: x(3)-D.

NAME: Glu / Leu / Phe / Val dehydrogenases active site.
CONSENSUS: [LIV]-x(2)-G-G-[SAG]-K-x-[GV]-x(3)-[DNST]-[PL].

NAME: D-amino acid oxidases signature.
CONSENSUS: [LIVM](2)-H-[NHA]-Y-G-x-[GSA](2)-x-G-x(5)-G-x-A.

NAME: Pyridoxamine 5'-phosphate oxidase signature.
CONSENSUS: [LIVF]-E-F-W-[QHG]-x(4)-R-[LIVM]-H-[DNE]-R.

NAME: Copper amine oxidase topaquinone signature.
CONSENSUS: [LIVM]-[LIVMA]-[LIVM]-x(4)-T-x(2)-N-Y-[DE]-[YN].

NAME: Copper amine oxidase copper-binding site signature.
CONSENSUS: T-x-G-x(2)-H-[LIVMF]-x(3)-E-[DE]-x-P.

NAME: Lysyl oxidase putative copper-binding region signature.
CONSENSUS: W-E-W-H-S-C-H-Q-H-Y-H.

NAME: Delta 1-pyrroline-5-carboxylate reductase signature.
CONSENSUS: [PALF]-x(2,3)-[LIV]-x(3)-[LIVM]-[STAC]-[STV]-x-[GAN]-G-x-T-x(2)-[AG]-
CONSENSUS: [LIV]-x(2)-[LMF]-[DENQK].

NAME: Dihydrofolate reductase signature.
CONSENSUS: [LVAGC]-[LIF]-G-x(4)-[LIVMF]-P-W-x(4,5)-[DE]-x(3)-[FYIV]-x(3)-[STIQ].

NAME: Tetrahydrofolate dehydrogenase/cyclohydrolase signature 1.
CONSENSUS: [EQ]-x-[EQK]-[LIVM](2)-x(2)-[LIVM]-x(2)-[LIVMY]-N-x-[DN]-x(5)-[LIVMF](3)-

CONSENSUS: Q-L-P-[LV].

NAME: Tetrahydrofolate dehydrogenase/cyclohydrolase signature 2.
CONSENSUS: P-G-G-V-G-P-[MF]-T-[IV].

NAME: Oxygen oxidoreductases covalent FAD-binding site.
CONSENSUS: P-x(10)-[DE]-[LIVM]-x(3)-[LIVM]-x(9)-[LIVM]-x(3)-[GSA]-[GST]-G-H.

NAME: Pyridine nucleotide-disulphide oxidoreductases class-I active site.
CONSENSUS: G-G-x-C-[LIVA]-x(2)-G-C-[LIVM]-P.

NAME: Pyridine nucleotide-disulphide oxidoreductases class-II active site.
CONSENSUS: C-x(2)-C-D-[GA]-x(2,4)-[FY]-x(4)-[LIVM]-x-[LIVM](2)-G(3)-[DN].

NAME: Respiratory-chain NADH dehydrogenase subunit 1 signature 1.
CONSENSUS: G-[LIVMFYKRS]-[LIVMAGP]-Q-x-[LIVMFY]-x-D-[AGIM]-[LIVMFTA]-K-[LVMYST]-
CONSENSUS: [LIVMFYG]-x-[KR]-[EQG].

NAME: Respiratory-chain NADH dehydrogenase subunit 1 signature 2.
CONSENSUS: P-F-D-[LIVMFYQ]-[STAGPVM]-E-[GAC]-E-x-[EQ]-[LIVMS]-x(2)-G.

NAME: Respiratory-chain NADH dehydrogenase 20 Kd subunit signature.
CONSENSUS: [GN]-x-D-[KRST]-[LIVMF](2)-P-[IV]-D-[LIVMFYW](2)-x-P-x-C-P-[PT].

NAME: Respiratory-chain NADH dehydrogenase 24 Kd subunit signature.
CONSENSUS: D-x(2)-F-[ST]-x(5)-C-L-G-x-C-x(2)-[GA]-P.

NAME: Respiratory chain NADH dehydrogenase 30 Kd subunit signature.
CONSENSUS: E-R-E-x(2)-[DE]-[LIVMF](2)-x(6)-[HK]-x(3)-[KRP]-x-[LIVM]-[LIVMS].

NAME: Respiratory chain NADH dehydrogenase 49 Kd subunit signature.
CONSENSUS: [LIVMH]-H-[RT]-[GA]-x-E-K-[LIVMT]-x-E-x-[KRQ].

NAME: Respiratory-chain NADH dehydrogenase 51 Kd subunit signature 1.
CONSENSUS: G-[AM]-G-[AR]-Y-[LIVM]-C-G-[DE](2)-[STA](2)-[LIM](2)-[EN]-S.

NAME: Respiratory-chain NADH dehydrogenase 51 Kd subunit signature 2.
CONSENSUS: E-S-C-G-x-C-x-P-C-R-x-G.

NAME: Respiratory-chain NADH dehydrogenase 75 Kd subunit signature 1.
CONSENSUS: P-x(2)-C-[YWS]-x(7)-G-x-C-R-x-C.

NAME: Respiratory-chain NADH dehydrogenase 75 Kd subunit signature 2.
CONSENSUS: C-P-x-C-[DE]-x-[GS](2)-x-C-x-L-Q.

NAME: Respiratory-chain NADH dehydrogenase 75 Kd subunit signature 3.
CONSENSUS: R-C-[LIVM]-x-C-x-R-C-[LIVM]-x-[FY].

NAME: Nitrite and sulfite reductases iron-sulfur/siroheme-binding site.
CONSENSUS: [STV]-G-C-x(3)-C-x(6)-[DE]-[LIVMF]-[GAT]-[LIVMF].

NAME: Uricase signature.
CONSENSUS: L-x-[LV]-L-K-[ST]-T-x-S-x-F-x(2)-[FY]-x(4)-[FY].

NAME: Heme-copper oxidase catalytic subunit, copper B binding region signature.
CONSENSUS: [YWG]-[LIVFYWTA](2)-[VGS]-H-[LNP]-x-V-x(44,47)-H-H.

NAME: CO II and nitrous oxide reductase dinuclear copper centers signature.
CONSENSUS: V-x-H-x(33,40)-C-x(3)-C-x(3)-H-x(2)-M.

NAME: Cytochrome c oxidase subunit Vb, zinc binding region signature.
CONSENSUS: [LIVM](2)-[FYW]-x(10)-C-x(2)-C-G-x(2)-[FY]-K-L.

NAME: Multicopper oxidases signature 1.
CONSENSUS: G-x-[FYW]-x-[LIVMFYW]-x-[CST]-x(8)-G-[LM]-x(3)-[LIVMFYW].

NAME: Multicopper oxidases signature 2.
CONSENSUS: H-C-H-x(3)-H-x(3)-[AG]-[LM].

NAME: Peroxidases proximal heme-ligand signature.
CONSENSUS: [DET]-[LIVMTA]-x(2)-[LIVM]-[LIVMSTAG]-[SAG]-[LIVMSTAG]-H-[STA]-[LIVMFY].

NAME: Peroxidases active site signature.
CONSENSUS: [SGATV]-x(3)-[LIVMA]-R-[LIVMA]-x-[FW]-H-x-[SAC].

NAME: Catalase proximal heme-ligand signature.
CONSENSUS: R-[LIVMFSTAN]-F-[GASTNP]-Y-x-D-[AST]-[QEH].

NAME: Catalase proximal active site signature.
CONSENSUS: [IF]-x-[RH]-x(4)-[EQ]-R-x(2)-H-x(2)-[GAS]-[GASTF]-[GAST].

NAME: Glutathione peroxidases selenocysteine active site.
CONSENSUS: [GN]-[RKHNFYC]-x-[LIVMFC]-[LIVMF](2)-x-N-[VT]-x-[STC]-x-C-[GA]-x-T.

NAME: Glutathione peroxidases signature 2.
CONSENSUS: [LIV]-[AGD]-F-P-[CS]-[NG]-Q-F.

NAME: Lipoxygenases iron-binding region signature 1.
CONSENSUS: H-[EQ]-x(3)-H-x-[LM]-[NQRC]-[GST]-H-[LIVMSTAC](3)-E.

NAME: Lipoxygenases iron-binding region signature 2.
CONSENSUS: [LIVMA]-H-P-[LIVM]-x-[KRQ]-[LIVMF](2)-x-[AP]-H.

NAME: Extradiol ring-cleavage dioxygenases signature.
CONSENSUS: [GNTIV]-x-H-x(5,7)-[LIVMF]-Y-x(2)-[DENTA]-P-x-[GP]-x(2,3)-E.

NAME: Intradiol ring-cleavage dioxygenases signature.
CONSENSUS: [LIVM]-x-G-x-[LIVM]-x(4)-[GS]-x(2)-[LIVM]-x(4)-[LIVM]-[DE]-[LIVMFY]-
CONSENSUS: x(6)-G-x-[FY].

NAME: Indoleamine 2,3-dioxygenase signature 1.
CONSENSUS: G-G-S-[AN]-[GA]-Q-S-S-x(2)-Q.

NAME: Indoleamine 2,3-dioxygenase signature 2.
CONSENSUS: [FY]-L-[DQ]-[DE]-[LIVM]-x(2)-Y-M-x(3)-H-[KR].

NAME: Bacterial ring hydroxylating dioxygenases alpha-subunit signature.
CONSENSUS: C-x-H-R-[GA]-x(8)-G-N-x(5)-C-x-[FY]-H.

NAME: Bacterial luciferase subunits signature.
CONSENSUS: [GA]-[LIVM]-P-[LIVM]-x-[LIVMFY]-x-W-x(6)-[RK]-x(6)-Y-x(3)-[AR].

NAME: ubiH/COQ6 monooxygenase family signature.
CONSENSUS: H-P-[LIV]-[AG]-G-Q-G-x-N-x-G-x(2)-D.

NAME: Biopterin-dependent aromatic amino acid hydroxylases signature.
CONSENSUS: P-D-x(2)-H-[DE]-[LI]-[LIVMF]-G-H-[LIVMC]-P.

NAME: Copper type II, ascorbate-dependent monooxygenases signature 1.
CONSENSUS: H-H-M-x(2)-F-x-C.

NAME: Copper type II, ascorbate-dependent monooxygenases signature 2.
CONSENSUS: H-x-F-x(4)-H-T-H-x(2)-G.

NAME: Tyrosinase CuA-binding region signature.
CONSENSUS: H-x(4,5)-F-[LIVMFTP]-x-[FW]-H-R-x(2)-[LM]-x(3)-E.

NAME: Tyrosinase and hemocyanins CuB-binding region signature.
CONSENSUS: D-P-x-F-[LIVMFYW]-x(2)-H-x(3)-D.

NAME: Fatty acid desaturases family 1 signature.
CONSENSUS: G-E-x-[FY]-H-N-[FY]-H-H-x-F-P-x-D-Y.

NAME: Fatty acid desaturases family 2 signature.
CONSENSUS: [ST]-[SA]-x(3)-[QR]-[LI]-x(5,6)-D-Y-x(2)-[LIVMFYW]-[LIVM]-[DE].

NAME: Cytochrome P450 cysteine heme-iron ligand signature.
CONSENSUS: [FW]-[SGNH]-x-[GD]-x-[RHPT]-x-C-[LIVMFAP]-[GAD].

NAME: Heme oxygenase signature.
CONSENSUS: L-L-V-A-H-A-Y-T-R.

NAME: Copper/Zinc superoxide dismutase signature 1.
CONSENSUS: [GA]-[IFAT]-H-[LIVF]-H-x(2)-[GP]-[SDG]-x-[STAGD].

NAME: Copper/Zinc superoxide dismutase signature 2.
CONSENSUS: G-[GN]-[SGA]-G-x-R-x-[SGA]-C-x(2)-[IV].

NAME: Manganese and iron superoxide dismutases signature.
CONSENSUS: D-x-W-E-H-[STA]-[FY](2).

NAME: Ribonucleotide reductase large subunit signature.
CONSENSUS: W-x(2)-[LF]-x(6,7)-G-[LIVM]-[FYRA]-[NH]-x(3)-[STAQLIVM]-[ASC]-x(2)-
CONSENSUS: [PA].

NAME: Ribonucleotide reductase small subunit signature.
CONSENSUS: [IVMSEQ]-E-x(1,2)-[LIVTA]-[HY]-[GSA]-x-[STAVM]-Y-x(2)-[LIVMQ]-x(3)-
CONSENSUS: [LIFY]-[IVFYCSA].

NAME: Nitrogenases component 1 alpha and beta subunits signature 1.
CONSENSUS: [LIVMFYH]-[LIVMFST]-H-[AG]-[AGSP]-[LIVMNQA]-[AG]-C.

NAME: Nitrogenases component 1 alpha and beta subunits signature 2.
CONSENSUS: [STANQ]-[ET]-C-x(5)-G-D-[DN]-[LIVMT]-x-[STAGR]-[LIVMFYST].

NAME: NifH/frxC family signature 1.
CONSENSUS: E-x-G-G-P-x(2)-[GA]-x-G-C-[AG]-G.

NAME: NifH/frxC family signature 2.
CONSENSUS: D-x-L-G-D-V-V-C-G-G-F-[AG]-x-P.

NAME: Nickel-dependent hydrogenases large subunit signature 1.
CONSENSUS: R-G-[LIVMF]-E-x(15)-[QESM]-R-x-C-G-[LIVM]-C.

NAME: Nickel-dependent hydrogenases large subunit signature 2.
CONSENSUS: [FY]-D-P-C-[LIM]-[ASG]-C-x(2,3)-H.

NAME: Glutamyl-tRNA reductase signature.
CONSENSUS: H-[LIVM]-x(2)-[LIVM]-[GSTAC](3)-[LIVM]-[DEQ]-S-[LIVMA]-[LIVM](2)-[GF]-E-
CONSENSUS: x-[QR]-[IV]-[LIT]-[STAG]-Q-[LIVM]-[KR].

NAME: Bacterial-type phytoene dehydrogenase signature.
CONSENSUS: [NG]-x-[FYWV]-[LIVMF]-x-G-[AGC]-[GS]-[TA]-[HQT]-P-G-[STAV]-G-[LIVM]-
CONSENSUS: x(5)-[GS].

NAME: Glycine radical signature.
CONSENSUS: [STIV]-x-R-[IVT]-[CSA]-G-Y-x-[GACV].

NAME: Ergosterol biosynthesis ERG4/ERG24 family signature 1.
CONSENSUS: G-x(2)-[LIVM]-Y-D-x-[FY]-x-G-x(2)-L-N-P-R.

NAME: Ergosterol biosynthesis ERG4/ERG24 family signature 2.
CONSENSUS: [LIVM](2)-H-R-x(2)-R-D-x(3)-C-x(2)-K-Y-G.

NAME: NNMT/PNMT/TEMT family of methyltransferases signature.
CONSENSUS: L-I-D-I-G-S-G-P-T-[IV]-Y-Q-L-L-S-A-C.

NAME: RNA methyltransferase trmA family signature 1.
CONSENSUS: [DN]-P-[PA]-R-x-G-x(14,16)-[LIVM](2)-Y-x-S-C-N-x(2)-T.

NAME: RNA methyltransferase trmA family signature 2.
CONSENSUS: [LIVMF]-D-x-F-P-[QHY]-[ST]-x-H-[LIVMFY]-E.

NAME: Thymidylate synthase active site.
CONSENSUS: R-x(2)-[LIVM]-x(3)-[FW]-[QN]-x(8,9)-[LV]-x-P-C-[HAVM]-x(3)-[QMT]-[FYW]-
CONSENSUS: x-[LV].

NAME: Ribosomal RNA adenine dimethylases signature.
CONSENSUS: [LIVM]-[LIVMFY]-[DE]-x-G-[STAPV]-G-x-[GA]-x-[LIVMF]-[ST]-x(2)-[LIVM]-
CONSENSUS: x(6)-[LIVMY]-x-[STAGV]-[LIVMFYHC]-E-x-D.

NAME: Methylated-DNA--protein-cysteine methyltransferase active site.
CONSENSUS: [LIVMF]-P-C-H-R-[LIVMF](2).

NAME: N-6 Adenine-specific DNA methylases signature.
CONSENSUS: [LIVMAC]-[LIVFYWA]-x-[DN]-P-P-[FYW].

NAME: N-4 cytosine-specific DNA methylases signature.
CONSENSUS: [LIVMF]-T-S-P-P-[FY].

NAME: C-5 cytosine-specific DNA methylases active site.
CONSENSUS: [DENKS]-x-[FLIV]-x(2)-[GSTC]-x-P-C-x(2)-[FYWLIM]-S.

NAME: C-5 cytosine-specific DNA methylases C-terminal signature.
CONSENSUS: [RKQGTF]-x(2)-G-N-[STAG]-[LIVMF]-x(3)-[LIVMT]-x(3)-[LIVM]-x(3)-[LIVM].

NAME: Protein-L-isoaspartate(D-aspartate) O-methyltransferase signature.
CONSENSUS: [GSA]-D-G-x(2)-G-[FYWV]-x(3)-[AS]-P-[FY]-[DN]-x-I.

NAME: Uroporphyrin-III C-methyltransferase signature 1.
CONSENSUS: [LIVM]-[GS]-[STAL]-G-P-G-x(3)-[LIVMFY]-[LIVM]-T-[LIVM]-[KRHQG]-[AG].

NAME: Uroporphyrin-III C-methyltransferase signature 2.
CONSENSUS: V-x(2)-[LI]-x(2)-G-D-x(3)-[FYW]-[GS]-x(8)-[LIVF]-x(5,6)-[LIVMFYWPAC]-
CONSENSUS: x-[LIVMY]-x-P-G.

NAME: ubiE/COQ5 methyltransferase family signature 1.
CONSENSUS: Y-D-x-M-N-x(2)-[LIVM]-S-x(3)-H-x(2)-W.

NAME: ubiE/COQ5 methyltransferase family signature 2.
CONSENSUS: R-V-[LIVM]-K-[PV]-G-G-x-[LIVMF]-x(2)-[LIVM]-E-x-S.

NAME: Serine hydroxymethyltransferase pyridoxal-phosphate attachment site.
CONSENSUS: [DEH]-[LIVMFY]-x-[STMV]-[GST]-[ST](2)-H-K-[ST]-[LF]-x-G-[PAC]-[RQ]-
CONSENSUS: [GSA]-[GA].

NAME: Phosphoribosylglycinamide formyltransferase active site.
CONSENSUS: G-x-[STM]-[IVT]-x-[FYWVQ]-[VMAT]-x-[DEVM]-x-[LIVMY]-D-x-G-x(2)-[LIVT]-
CONSENSUS: x(6)-[LIVM].

NAME: Aspartate and ornithine carbamoyltransferases signature.
CONSENSUS: F-x-[EK]-x-S-[GT]-R-T.

NAME: Transketolase signature 1.
CONSENSUS: R-x(3)-[LIVMTA]-[DENQSTHKF]-x(5,6)-[GSN]-G-H-[PLIVMF]-[GSTA]-x(2)-
CONSENSUS: [LIMC]-[GS].

NAME: Transketolase signature 2.
CONSENSUS: G-[DEQGSA]-[DN]-G-[PAEQ]-[ST]-[HQ]-x-[PAGM]-[LIVMYAC]-[DEFYW]-x(2)-
CONSENSUS: [STAP]-x(2)-[RGA].

NAME: Transaldolase signature 1.
CONSENSUS: [DG]-[IVSA]-T-[ST]-N-P-[STA]-[LIVMF](2).

NAME: Transaldolase active site.
CONSENSUS: [LIVM]-x-[LIVM]-K-[LIVM]-[PAS]-x-[ST]-x-[DENQPAS]-G-[LIVM]-x-[AGV]-x-
CONSENSUS: [QEKRST]-x-[LIVM].

NAME: Acyltransferases ChoActase / COT / CPT family signature 1.
CONSENSUS: [LI]-P-x-[LVP]-P-[IVTA]-P-x-[LIVM]-x-[DENQAS]-[ST]-[LIVM]-x(2)-[LY].

NAME: Acyltransferases ChoActase / COT / CPT family signature 2.
CONSENSUS: R-[FYW]-x-[DA]-[KA]-x(0,1)-[LIVMFY]-x-[LIVMFY](2)-x(3)-[DNS]-[GSA]-x(6)-
CONSENSUS: [DE]-[HS]-x(3)-[DE]-[GA].

NAME: Thiolases acyl-enzyme intermediate signature.
CONSENSUS: [LIVM]-[NST]-x(2)-C-[SAGLI]-[ST]-[SAG]-[LIVMFYNS]-x-[STAG]-[LIVM]-x(6)-
CONSENSUS: [LIVM].

NAME: Thiolases signature 2.
CONSENSUS: N-x(2)-G-G-x-[LIVM]-[SA]-x-G-H-P-x-G-x-[ST]-G.

NAME: Thiolases active site.
CONSENSUS: [AG]-[LIVMA]-[STAGLIVM]-[STAG]-[LIVMA]-C-x-[AG]-x-[AG]-x-[AG]-x-[SAG].

NAME: Chloramphenicol acetyltransferase active site.
CONSENSUS: Q-[LIV]-H-H-[SA]-x(2)-D-G-[FY]-H.

NAME: Hexapeptide-repeat containing-transferases signature.
CONSENSUS: [LIV]-[GAED]-x(2)-[STAV]-x-[LIV]-x(3)-[LIVAC]-x-[LIV]-[GAED]-x(2)-
CONSENSUS: [STAVR]-x-[LIV]-[GAED]-x(2)-[STAV]-x-[LIV]-x(3)-[LIV].

NAME: Beta-ketoacyl synthases active site.
CONSENSUS: G-x(4)-[LIVMFAP]-x(2)-[AGC]-C-[STA](2)-[STAG]-x(3)-[LIVMF].

NAME: Chalcone and stilbene synthases active site.

CONSENSUS: R-[LIVMFYS]-x-[LIVM]-x-[QHG]-x-G-C-[FYNA]-[GA]-G-[GA]-[STAV]-x-[LIVMF]-
CONSENSUS: [RA].

NAME: Myristoyl-CoA:protein N-myristoyltransferase signature 1.
CONSENSUS: E-I-N-F-L-C-x-H-K.

NAME: Myristoyl-CoA:protein N-myristoyltransferase signature 2.
CONSENSUS: K-F-G-x-G-D-G.

NAME: Gamma-glutamyltranspeptidase signature.
CONSENSUS: T-[STA]-H-x-[ST]-[LIVMA]-x(4)-G-[SN]-x-V-[STA]-x-T-x-T-[LIVM]-[NE]-
CONSENSUS: x(1,2)-[FY]-G.

NAME: Transglutaminases active site.
CONSENSUS: [GT]-Q-[CA]-W-V-x-[SA]-[GA]-[IVT]-x(2)-T-x-[LMSC]-R-[CSA]-[LV]-G.

NAME: Phosphorylase pyridoxal-phosphate attachment site.
CONSENSUS: E-A-[SC]-G-x-[GS]-x-M-K-x(2)-[LM]-N.

NAME: UDP-glycosyltransferases signature.
CONSENSUS: [FW]-x(2)-Q-x(2)-[LIVMYA]-[LIMV]-x(4,6)-[LVGAC]-[LVFYA]-[LIVMF]-[STAGCM]-
CONSENSUS: [HNQ]-[STAGC]-G-x(2)-[STAG]-x(3)-[STAGL]-[LIVMFA]-x(4)-[PQR]-[LIVMT]-
CONSENSUS: x(3)-[PA]-x(3)-[DES]-[QEHN].

NAME: Purine/pyrimidine phosphoribosyl transferases signature.
CONSENSUS: [LIVMFYWCTA]-[LIVM]-[LIVMA]-[LIVMFC]-[DE]-D-[LIVMS]-[LIVM]-[STAVD]-
CONSENSUS: [STAR]-[GAC]-x-[STAR].

NAME: Glutamine amidotransferases class-I active site.
CONSENSUS: [PAS]-[LIVMFYT]-[LIVMFY]-G-[LIVMFY]-C-[LIVMFYN]-G-x-[QEH]-x-[LIVMFA].

NAME: Glutamine amidotransferases class-II active site.
CONSENSUS: <x(0,11)-C-[GS]-[IV]-[LIVMFYW]-[AG].

NAME: Purine and other phosphorylases family 1 signature.
CONSENSUS: [GST]-x-G-[LIVM]-G-x-[PA]-S-x-[GSTA]-I-x(3)-E-L.

NAME: Purine and other phosphorylases family 2 signature.
CONSENSUS: [LIV]-x(3)-G-x(2)-H-x-[LIVMFY]-x(4)-[LIVMF]-x(3)-[ATV]-x(1,2)-[LIVM]-x-
CONSENSUS: [ATV]-x(4)-[GN]-x(3,4)-[LIVMF](2)-x(2)-[STN]-[SA]-x-G-[GS]-[LIVM].

NAME: Thymidine and pyrimidine-nucleoside phosphorylases signature.
CONSENSUS: S-[GS]-R-[GA]-[LIV]-x(2)-[TA]-[GA]-G-T-x-D-x-[LIV]-E.

NAME: ATP phosphoribosyltransferase signature.
CONSENSUS: E-x(5)-G-x-[SAG]-x(2)-[IV]-x-D-[LIV]-x(2)-[ST]-G-x-T-[LM].

NAME: NAD:arginine ADP-ribosyltransferases signature.
CONSENSUS: [FY]-x-[FY]-K-x(2)-H-[FY]-x-L-[ST]-x-A.

NAME: Prolipoprotein diacylglyceryl transferase signature.
CONSENSUS: G-R-x-[GA]-N-F-[LIVMF]-N-x-E-x(2)-G.

NAME: S-adenosylmethionine synthetase signature 1.
CONSENSUS: G-A-G-D-Q-G-x(3)-G-Y.

NAME: S-adenosylmethionine synthetase signature 2.
CONSENSUS: G-[GA]-G-[ASC]-F-S-x-K-[DE].

NAME: Polyprenyl synthetases signature 1.
CONSENSUS: [LIVM](2)-x-D-D-x(2,4)-D-x(4)-R-R-[GH].

NAME: Polyprenyl synthetases signature 2.
CONSENSUS: [LIVMFY]-G-x(2)-[FYL]-Q-[LIVM]-x-D-D-[LIVMFY]-x-[DNG].

NAME: Squalene and phytoene synthases signature 1.
CONSENSUS: Y-[CSAM]-x(2)-[VSG]-A-[GSA]-[LIVAT]-[IV]-G-x(2)-[LMSC]-x(2)-[LIV].

NAME: Squalene and phytoene synthases signature 2.
CONSENSUS: [LIVM]-G-x(3)-Q-x(2,3)-N-[IF]-x-R-D-[LIVMFY]-x(2)-[DE]-x(4,7)-R-x-[FY]-
CONSENSUS: x-P.

NAME: Protein prenyltransferases alpha subunit repeat signature.
CONSENSUS: [PSIAV]-x-[NDFV]-[NEQIY]-x-[LIVMAGP]-W-[NQSTHF]-[FYHQ]-[LIVMR].

NAME: Riboflavin synthase alpha chain family signature.
CONSENSUS: [LIVMF]-x(5)-G-[STADNQ]-[KREQIYW]-V-N-[LIVM]-E.

NAME: Dihydropteroate synthase signature 1.
CONSENSUS: [LIVM]-x-[AG]-[LIVMF](2)-N-x-T-x-D-S-F-x-D-x-[SG].

NAME: Dihydropteroate synthase signature 2.
CONSENSUS: [GE]-[SA]-x-[LIVM](2)-D-[LIVM]-G-[GP]-x(2)-[STA]-x-P.

NAME: EPSP synthase signature 1.
CONSENSUS: [LIVM]-x(2)-[GN]-N-[SA]-G-T-[STA]-x-R-x-[LIVMY]-x-[GSTA].

NAME: EPSP synthase signature 2.
CONSENSUS: [KR]-x-[KH]-E-[CST]-[DNE]-R-[LIVM]-x-[STA]-[LIVMC]-x(2)-[EN]-[LIVMF]-x-
CONSENSUS: [KRA]-[LIVMF]-G.

NAME: FLAP/GST2/LTC4S family signature.
CONSENSUS: G-x(3)-F-E-R-V-[FY]-x-A-[NQ]-x-N-C.

NAME: Aminotransferases class-I pyridoxal-phosphate attachment site.
CONSENSUS: [GS]-[LIVMFYTAC]-[GSTA]-K-x(2)-[GSALVN]-[LIVMFA]-x-[GNAR]-x-R-[LIVMA]-
CONSENSUS: [GA].

NAME: Aminotransferases class-II pyridoxal-phosphate attachment site.
CONSENSUS: T-[LIVMFYW]-[STAG]-K-[SAG]-[LIVMFYWR]-[SAG]-x(2)-[SAG].

NAME: Aminotransferases class-III pyridoxal-phosphate attachment site.
CONSENSUS: [LIVMFYWC](2)-x-D-E-[LIVMA]-x(2)-[GP]-x(0,1)-[LIVMFYWAG]-x(0,1)-[SACR]-x-
CONSENSUS: [GSAD]-x(12,16)-D-[LIVMFYWC]-x(2,3)-[GSA]-K-x(3)-[GSTADN]-[GSA].

NAME: Aminotransferases class-IV signature.
CONSENSUS: E-x-[STAGCI]-x(2)-N-[LIVMFAC]-[FY]-x(6,12)-[LIVMF]-x-T-x(6,8)-[LIVM]-x-
CONSENSUS: [GS]-[LIVM]-x-[KR].

NAME: Aminotransferases class-V pyridoxal-phosphate attachment site.
CONSENSUS: [LIVFYCHT]-[DGH]-[LIVMFYAC]-[LIVMFYA]-x(2)-[GSTAC]-[GSTA]-[HQR]-K-
CONSENSUS: x(4,6)-G-x-[GSAT]-x-[LIVMFYSAC].

NAME: Hexokinases signature.
CONSENSUS: [LIVM]-G-F-[TN]-F-S-[FY]-P-x(5)-[LIVM]-[DNST]-x(3)-[LIVM]-x(2)-W-T-K-x-
CONSENSUS: [LF].

NAME: Galactokinase signature.
CONSENSUS: G-R-x-N-[LIV]-I-G-E-H-x-D-Y.

NAME: GHMP kinases putative ATP-binding domain.
CONSENSUS: [LIVM]-[PK]-x-[GSTA]-x(0,1)-G-L-[GS]-S-S-[GSA]-[GSTAC].

NAME: Phosphofructokinase signature.
CONSENSUS: [RK]-x(4)-G-H-x-Q-[QR]-G-G-x(5)-D-R.

NAME: pfkB family of carbohydrate kinases signature 1.
CONSENSUS: [AG]-G-x(0,1)-[GAP]-x-N-x-[STA]-x(6)-[GS]-x(9)-G.

NAME: pfkB family of carbohydrate kinases signature 2.
CONSENSUS: [DNSK]-[PSTV]-x-[SAG](2)-[GD]-D-x(3)-[SAGV]-[AG]-[LIVMFY]-[LIVMSTAP].

NAME: ROK family signature.
CONSENSUS: [LIVM]-x(2)-G-[LIVMFCT]-G-x-[GA]-[LIVMFA]-x(8)-G-x(3,5)-[GATP]-x(2)-
CONSENSUS: G-[RKH].

NAME: Phosphoribulokinase signature.
CONSENSUS: K-[LIVM]-x-R-D-x(3)-R-G-x-[ST]-x-E.

NAME: Thymidine kinase cellular-type signature.
CONSENSUS: [GA]-x(1,2)-[DE]-x-Y-x-[STAP]-x-C-[NKR]-x-[CH]-[LIVMFYWH].

NAME: FGGY family of carbohydrate kinases signature 1.
CONSENSUS: [MFYGS]-x-[PST]-x(2)-K-[LIVMFYW]-x-W-[LIVMF]-x-[DENQTKR]-[ENQH].

NAME: FGGY family of carbohydrate kinases signature 2.
CONSENSUS: [GSA]-x-[LIVMFYW]-x-G-[LIVM]-x(7,8)-[HDENQ]-[LIVMF]-x(2)-[AS]-[STAIVM]-
CONSENSUS: [LIVMFY]-[DEQ].

NAME: Protein kinases ATP-binding region signature.
CONSENSUS: [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-
CONSENSUS: x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K.

NAME: Serine/Threonine protein kinases active-site signature.
CONSENSUS: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3).

NAME: Tyrosine protein kinases specific active-site signature.
CONSENSUS: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC](3).

NAME: Protein kinase domain profile.

NAME: Casein kinase II regulatory subunit signature.
CONSENSUS: C-P-x-[LIVMY]-x-C-x(5)-L-P-[LIVMC]-G-x(9)-V-[KR]-x(2)-C-P-x-C.

NAME: Pyruvate kinase active site signature.
CONSENSUS: [LIVAC]-x-[LIVM](2)-[SAPCV]-K-[LIV]-E-[NKRST]-x-[DEQH]-[GSTA]-[LIVM].

NAME: Shikimate kinase signature.
CONSENSUS: [KR]-x(2)-E-x(3)-[LIVMF]-x(8,12)-[LIVMF](2)-[SA]-x-G(3)-x-[LIVMF].

NAME: Prokaryotic diacylglycerol kinase signature.
CONSENSUS: E-x-[LIVM]-N-[ST]-[SA]-[LIV]-E-x(2)-V-D.

NAME: Phosphatidylinositol 3- and 4-kinases signature 1.
CONSENSUS: [LIVMFAC]-K-x(1,3)-[DEA]-[DE]-[LIVMC]-R-Q-[DE]-x(4)-Q.

NAME: Phosphatidylinositol 3- and 4-kinases signature 2.
CONSENSUS: [GS]-x-[AV]-x(3)-[LIVM]-x(2)-[FYH]-[LIVM](2)-x-[LIVMF]-x-D-R-H-x(2)-N.

NAME: Acetate and butyrate kinases family signature 1.
CONSENSUS: [LIVM](2)-x-[LIVM]-N-x-G-S-[ST]-S-x-[KE].

NAME: Acetate and butyrate kinases family signature 2.
CONSENSUS: [LIVMA](2)-x(2)-H-x-G-x-G-x-[ST]-[LIVM]-x-[AV]-x(3)-G.

NAME: Phosphoglycerate kinase signature.
CONSENSUS: [KRHGTCV]-[VT]-[LIVMF]-[LIVMC]-R-x-D-x-N-[SACV]-P.

NAME: Aspartokinase signature.
CONSENSUS: [LIVM]-x-K-[FY]-G-G-[ST]-[SC]-[LIVM].

NAME: Glutamate 5-kinase signature.
CONSENSUS: [GSTN]-x(2)-G-x-G-[GC]-[IM]-x-[STA]-K-[LIVM]-x-[SA]-[TCA]-x(2)-[GALV]-
CONSENSUS: x(3)-G.

NAME: ATP:guanido phosphotransferases active site.
CONSENSUS: C-P-x(0,1)-[ST]-N-[IL]-G-T.

NAME: PTS HPR component histidine phosphorylation site signature.
CONSENSUS: G-[LIVM]-H-[STA]-R-[PA]-[GSTA]-[STAM].

NAME: PTS HPR component serine phosphorylation site signature.
CONSENSUS: [GSADE]-[KREQTV]-x(4)-[KRN]-S-[LIVMF](2)-x-[LIVM]-x(2)-[LIVM]-[GAD].

NAME: PTS EIIA domains phosphorylation site signature 1.
CONSENSUS: G-x(2)-[LIVMF](3)-H-[LIVMF]-G-[LIVMF]-x-T-[ALV].

NAME: PTS EIIA domains phosphorylation site signature 2.
CONSENSUS: [DENQ]-x(6)-[LIVMF]-[GA]-x(2)-[LIVM]-A-[LIVM]-P-H-[GAC].

NAME: PTS EIIB domains cysteine phosphorylation site signature.
CONSENSUS: N-[LIVMFY]-x(5)-C-x-T-R-[LIVMF]-x-[LIVMF]-x-[LIVM]-x-[DQ].

NAME: Adenylate kinase signature.
CONSENSUS: [LIVMFYW](3)-D-G-[FYI]-P-R-x(3)-[NQ].

NAME: Nucleoside diphosphate kinases active site.
CONSENSUS: N-x(2)-H-[GA]-S-D-[SA]-[LIVMPKNE].

NAME: Guanylate kinase signature.
CONSENSUS: T-[ST]-R-x(2)-[KR]-x(2)-[DE]-x(2)-G-x(2)-Y-x-[FY]-[LIVMK].

NAME: Guanylate kinase domain profile.

NAME: Phosphoribosyl pyrophosphate synthetase signature.
CONSENSUS: D-[LI]-H-[SA]-x-Q-[IMST]-[QM]-G-[FY]-F-x(2)-P-[LIVMFC]-D.

NAME: 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase signature.
CONSENSUS: G-[PE]-R-x(2)-D-L-D-[LIVM](2).

NAME: Bacteriophage-type RNA polymerase family active site signature 1.
CONSENSUS: P-[LIVM]-x(2)-D-[GA]-[ST]-[AC]-[SN]-[GA]-[LIVMFY]-Q.

NAME: Bacteriophage-type RNA polymerase family active site signature 2.
CONSENSUS: [LIVMF]-x-R-x(3)-K-x(2)-[LIVMF]-M-[PT]-x(2)-Y.

NAME: Eukaryotic RNA polymerase II heptapeptide repeat.
CONSENSUS: Y-[ST]-P-[ST]-S-P-[STANK].

NAME: RNA polymerases beta chain signature.
CONSENSUS: G-x-K-[LIVMFA]-[STAC]-[GSTN]-x-[HSTA]-[GS]-[QNH]-K-G-[IVT].

NAME: RNA polymerases M / 15 Kd subunits signature.
CONSENSUS: F-C-x-[DEKST]-C-[GNK]-[DNSA]-[LIVMH]-[LIVM]-x(8,14)-C-x(2)-C.

NAME: RNA polymerases D / 30 to 40 Kd subunits signature.
CONSENSUS: N-[SGA]-[LIVMF]-R-R-x(9)-[SA]-x(3)-V-x(4)-N-x-[STA]-x(3)-[DN]-E-x-[LI]-
CONSENSUS: [GA]-x-R-[LI]-[GA]-[LIVM](2)-P.

NAME: RNA polymerases H / 23 Kd subunits signature.
CONSENSUS: H-[NEI]-[LIVM]-V-P-x-H-x(2)-[LIVM]-x(2)-[DE].

NAME: RNA polymerases K / 14 to 18 Kd subunits signature.
CONSENSUS: [ST]-x-[FY]-E-x-[AT]-R-x-[LIVM]-[GSA]-x-R-[SA]-x-Q.

NAME: RNA polymerases L / 13 to 16 Kd subunits signature.
CONSENSUS: [DE](2)-H-[ST]-[LIVM]-[GAP]-N-x(11)-V-x-[FM]-x(2)-Y-x(3)-H-P.

NAME: RNA polymerases N / 8 Kd subunits signature.
CONSENSUS: [LIVMF](2)-P-[LIVM]-x-C-F-[ST]-C-G.

NAME: DNA polymerase family A signature.
CONSENSUS: R-x(2)-[GSAV]-K-x(3)-[LIVMFY]-[AGQ]-x(2)-Y-x(2)-[GS]-x(3)-[LIVMA].

NAME: DNA polymerase family B signature.
CONSENSUS: [YA]-[GLIVMSTAC]-D-T-D-[SG]-[LIVMFTC]-x-[LIVMSTAC].

NAME: DNA polymerase family X signature.
CONSENSUS: G-[SG]-[LFY]-x-R-[GE]-x(3)-[SGCL]-x-D-[LIVM]-D-[LIVMFY](3)-x(2)-[SAP].

NAME: Galactose-1-phosphate uridyl transferase family 1 active site signature.
CONSENSUS: F-E-N-[RK]-G-x(3)-G-x(4)-H-P-H-x-Q.

NAME: Galactose-1-phosphate uridyl transferase family 2 signature.
CONSENSUS: D-L-P-I-V-G-G-[ST]-[LIVM](2)-[SA]-H-[DEN]-H-[FY]-Q-G-G.

NAME: ADP-glucose pyrophosphorylase signature 1.
CONSENSUS: [AG]-G-G-x-G-[STK]-x-L-x(2)-L-[TA]-x(3)-A-x-P-A-[LV].

NAME: ADP-glucose pyrophosphorylase signature 2.
CONSENSUS: W-[FY]-x-G-[ST]-A-[DNSH]-[AS]-[LIVMFYW].

NAME: ADP-glucose pyrophosphorylase signature 3.
CONSENSUS: [APV]-[GS]-M-G-[LIVMN]-Y-[IVC]-[LIVMFY]-x(2)-[DENPHK].

NAME: Phosphatidate cytidylyltransferase signature.
CONSENSUS: S-x-[LIVMF]-K-R-x(4)-K-D-x-[GSA]-x(2)-[LI]-[PG]-x-H-G-G-[LIVM]-x-D-R-
CONSENSUS: [LIVMFT]-D.

NAME: Ribonuclease PH signature.
CONSENSUS: C-[DE]-[LIVM](2)-Q-[GTA]-D-G-[SG]-x(2)-[TA]-A.

NAME: 2'-5'-oligoadenylate synthetases signature 1.
CONSENSUS: G-G-S-x-[AG]-[KR]-x-T-x-L-[KR]-[GST]-x-S-D-[AG].

NAME: 2'-5'-oligoadenylate synthetases signature 2.

CONSENSUS: R-P-V-I-L-D-P-x-[DE]-P-T.

NAME: CDP-alcohol phosphatidyltransferases signature.
CONSENSUS: D-G-x(2)-A-R-x(8)-G-x(3)-D-x(3)-D.

NAME: PEP-utilizing enzymes phosphorylation site signature.
CONSENSUS: G-[GA]-x-[TN]-x-H-[STA]-[STAV]-[LIVM](2)-[STAV]-[RG].

NAME: PEP-utilizing enzymes signature 2.
CONSENSUS: [DEQS]-x-[LIVMF]-S-[LIVMF]-G-[ST]-N-D-[LIVM]-x-Q-[LIVMFYGT]-[STALIV]-
CONSENSUS: [LIVMF]-[GAS]-x(2)-R.

NAME: Rhodanese signature 1.
CONSENSUS: [FY]-x(3)-H-[LIV]-P-G-A-x(2)-[LIVF].

NAME: Rhodanese C-terminal signature.
CONSENSUS: [AV]-x(2)-[FY]-[DEAP]-G-[GSA]-[WF]-x-E-[FYW].

NAME: CoA transferases signature 1.
CONSENSUS: [DN]-[GN]-x(2)-[LIVMFA](3)-G-G-F-x(3)-G-x-P.

NAME: CoA transferases signature 2.
CONSENSUS: [LF]-[HQ]-S-E-N-G-[LIVF](2)-[GA].

NAME: Phospholipase A2 histidine active site.
CONSENSUS: C-C-x(2)-H-x(2)-C.

NAME: Phospholipase A2 aspartic acid active site.
CONSENSUS: [LIVMA]-C-{LIVMFYWPCST}-C-D-x(5)-C.

NAME: Lipases, serine active site.
CONSENSUS: [LIV]-x-[LIVFY]-[LIVMST]-G-[HYWV]-S-x-G-[GSTAC].

NAME: Colipase signature.
CONSENSUS: Y-x(2)-Y-Y-x-C-x-C.

NAME: Lipolytic enzymes "G-D-S-L" family, serine active site.
CONSENSUS: [LIVMFYAG](4)-G-D-S-[LIVM]-x(1,2)-[TAG]-G.

NAME: Lipolytic enzymes "G-D-X-G" family, putative histidine active site.
CONSENSUS: [LIVMF](2)-x-[LIVMF]-H-G-G-[SAG]-[FY]-x(3)-[STDN]-x(2)-[ST]-H.

NAME: Lipolytic enzymes "G-D-X-G" family, putative serine active site.
CONSENSUS: [LIVM]-x-[LIVMF]-[SA]-G-D-S-[CA]-G-[GA]-x-L-[CA].

NAME: Carboxylesterases type-B serine active site.
CONSENSUS: F-[GR]-G-x(4)-[LIVM]-x-[LIV]-x-G-x-S-[STAG]-G.

NAME: Carboxylesterases type-B signature 2.
CONSENSUS: [ED]-D-C-L-[YT]-[LIV]-[DNS]-[LIV]-[LIVFYW]-x-[PQR].

NAME: Pectinesterase signature 1.
CONSENSUS: [GSTN]-x(5)-[LIVM]-x-[LIVM]-x(2)-G-x-Y-[DNK]-E-x-[LIVM]-x-[LIVM].

NAME: Pectinesterase signature 2.
CONSENSUS: G-[STAD]-[LIVMT]-D-F-I-F-G.

NAME: Peptidyl-tRNA hydrolase signature 1.
CONSENSUS: [FY]-x(2)-T-R-H-N-x-G-x(2)-[LIVMFA](2)-[DE].

NAME: Peptidyl-tRNA hydrolase signature 2.
CONSENSUS: [GS]-x(3)-H-N-G-[LIVM]-[KR]-[DNS]-[LIVMT].

NAME: Alkaline phosphatase active site.
CONSENSUS: [IV]-x-D-S-[GAS]-[GASC]-[GAST]-[GA]-T.

NAME: Histidine acid phosphatases phosphohistidine signature.
CONSENSUS: [LIVM]-x(2)-[LIVMA]-x(2)-[LIVM]-x-R-H-[GN]-x-R-x-[PAS].

NAME: Histidine acid phosphatases active site signature.
CONSENSUS: [LIVMF]-x-[LIVMFAG]-x(2)-[STAGI]-H-D-[STANQ]-x-[LIVM]-x(2)-[LIVMFY]-x(2)-
CONSENSUS: [STA].

NAME: Class A bacterial acid phosphatases signature.

CONSENSUS: G-S-Y-P-S-G-H-T.

NAME: 5'-nucleotidase signature 1.
CONSENSUS: [LIVM]-x-[LIVM](2)-[HEA]-[TI]-x-D-x-H-[GSA]-x-[LIVMF].

NAME: 5'-nucleotidase signature 2.
CONSENSUS: [FYP]-x(4)-[LIVM]-G-N-H-E-F-[DN].

NAME: Fructose-1-6-bisphosphatase active site.
CONSENSUS: [AG]-[RK]-L-x(1,2)-[LIV]-[FY]-E-x(2)-P-[LIVM]-[GSA].

NAME: Serine/threonine specific protein phosphatases signature.
CONSENSUS: [LIVM]-R-G-N-H-E.

NAME: Protein phosphatase 2A regulatory subunit PR55 signature 1.
CONSENSUS: E-F-D-Y-L-K-S-L-E-I-E-E-K-I-N.

NAME: Protein phosphatase 2A regulatory subunit PR55 signature 2.
CONSENSUS: N-[AG]-H-[TA]-Y-H-I-N-S-I-S-[LIVM]-N-S-D.

NAME: Protein phosphatase 2C signature.
CONSENSUS: [LIVMFY]-[LIVMFYA]-[GSAC]-[LIVM]-[FYC]-D-G-H-[GAV].

NAME: Tyrosine specific protein phosphatases active site.
CONSENSUS: [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY].

NAME: Tyrosine specific protein phosphatases profile.

NAME: Dual specificity protein phosphatase profile.

NAME: PTP type protein phosphatase profile.

NAME: Inositol monophosphatase family signature 1.
CONSENSUS: [FWV]-x(0,1)-[LIVM]-D-P-[LIVM]-D-[SG]-[ST]-x(2)-[FY]-x-[HKRNSTY].

NAME: Inositol monophosphatase family signature 2.
CONSENSUS: [WV]-D-x-[AC]-[GSA]-[GSAPV]-x-[LIVACP]-[LIV]-[LIVAC]-x(3)-[GH]-[GA].

NAME: Prokaryotic zinc-dependent phospholipase C signature.
CONSENSUS: H-Y-x-[GT]-D-[LIVM]-[DNS]-x-P-x-H-[PA]-x-N.

NAME: Phosphatidylinositol-specific phospholipase X-box domain profile.

NAME: Phosphatidylinositol-specific phospholipase Y-box domain profile.

NAME: 3'5'-cyclic nucleotide phosphodiesterases signature.
CONSENSUS: H-D-[LIVMFY]-x-H-x-[AG]-x(2)-[NQ]-x-[LIVMFY].

NAME: cAMP phosphodiesterases class-II signature.
CONSENSUS: H-x-H-L-D-H-[LIVM]-x-[GS]-[LIVMA]-[LIVM](2)-x-S-[AP].

NAME: Sulfatases signature 1.
CONSENSUS: [SAP]-[LIVMST]-[CS]-[STAC]-P-[STA]-R-x(2)-[LIVMFW](2)-[TR]-G.

NAME: Sulfatases signature 2.
CONSENSUS: G-[YV]-x-[ST]-x(2)-[IVA]-G-K-x(0,1)-[FYWK]-[HL].

NAME: AP endonucleases family 1 signature 1.
CONSENSUS: [APF]-D-[LIVMF](2)-x-[LIVM]-Q-E-x-K.

NAME: AP endonucleases family 1 signature 2.
CONSENSUS: D-[ST]-[FY]-R-[KH]-x(7,8)-[FYW]-[ST]-[FYW](2).

NAME: AP endonucleases family 1 signature 3.
CONSENSUS: N-x-G-x-R-[LIVM]-D-[LIVMFYH]-x-[LV]-x-S.

NAME: AP endonucleases family 2 signature 1.
CONSENSUS: H-x(2)-Y-[LIVMF]-[IM]-N-[LIVMCA]-[AG].

NAME: AP endonucleases family 2 signature 2.
CONSENSUS: [GR]-[LIVMF]-C-[LIVM]-D-T-C-H.

NAME: AP endonucleases family 2 signature 3.
CONSENSUS: [LIVMW]-H-x-N-[DE]-[SA]-K-x(3)-G-[SA]-x(2)-D.

NAME: Deoxyribonuclease I signature 1.
CONSENSUS: [LIVM](2)-[AP]-L-H-[STA](2)-P-x(5)-E-[LIVM]-[DN]-x-L-x-[DE]-V.

NAME: Deoxyribonuclease I signature 2.
CONSENSUS: G-D-F-N-A-x-C-[SA].

NAME: Endonuclease III iron-sulfur binding region signature.
CONSENSUS: C-x(3)-[KRS]-P-[KRAGL]-C-x(2)-C-x(5)-C.

NAME: Endonuclease III family signature.
CONSENSUS: [GST]-x-[LIVMF]-P-x(5)-[LIVMW]-x(2,3)-[LI]-[PAS]-G-V-[GA]-x(3)-[GAC]-
CONSENSUS: x(3)-[LIVM]-x(2)-[SALV]-[LIVMFYW]-[GANK].

NAME: Ribonuclease II family signature.
CONSENSUS: [HI]-[FYE]-[GSTAM]-[LIVM]-x(4,5)-Y-[STAL]-x-[FWVAC]-[TV]-[SA]-P-[LIVMA]-
CONSENSUS: [RQ]-[KR]-[FY]-x-D-x(3)-[HQ].

NAME: Ribonuclease III family signature.
CONSENSUS: [DEQ]-[RQ]-[LM]-E-[FYW]-[LV]-G-D-[SAR].

NAME: Bacterial Ribonuclease P protein component signature.
CONSENSUS: [LIVMFYS]-x(2)-A-x(2)-R-[NH]-[KRQL]-[LIVM]-[KRA]-R-x-[LIVMTA]-[KR].

NAME: Ribonuclease T2 family histidine active site 1.
CONSENSUS: [FYWL]-x-[LIVM]-H-G-L-W-P.

NAME: Ribonuclease T2 family histidine active site 2.
CONSENSUS: [LIVMF]-x(2)-[HDGTY]-[EQ]-[FYW]-x-[KR]-H-G-x-C.

NAME: Pancreatic ribonuclease family signature.
CONSENSUS: C-K-x(2)-N-T-F.

NAME: DNA/RNA non-specific endonucleases active site.
CONSENSUS: D-R-G-H-[QIL]-x(3)-A.

NAME: Thermonuclease family signature 1.
CONSENSUS: D-G-D-T-[LIVM]-x-[LIVMC]-x(9,10)-R-[LIVM]-x(2)-[LIVM]-D-x-P-E.

NAME: Thermonuclease family signature 2.
CONSENSUS: D-[KR]-Y-[GQ]-R-x-[LV]-[GA]-x-[IV]-[FYW].

NAME: Beta-amylase active site 1.
CONSENSUS: H-x-C-G-G-N-V-G-D.

NAME: Beta-amylase active site 2.
CONSENSUS: G-x-[SA]-G-E-[LIVM]-R-Y-P-S-Y.

NAME: Glucoamylase active site region signature.
CONSENSUS: [STN]-[GP]-x(1,2)-[DE]-x-W-E-E-x(2)-[GS].

NAME: Polygalacturonase active site.
CONSENSUS: [GSDENKRH]-x(2)-[VMFC]-x(2)-[GS]-H-G-[LIVMAG]-x(1,2)-[LIVM]-G-S.

NAME: Clostridium cellulosome enzymes repeated domain signature.
CONSENSUS: D-[LIVMFY]-[DNV]-x-[DNS]-x(2)-[LIVM]-[DN]-[SALM]-x-D-x(3)-[LIVMF]-x-
CONSENSUS: [RKS]-x-[LIVMF].

NAME: Chitinases family 18 active site.
CONSENSUS: [LIVMFY]-[DN]-G-[LIVMF]-[DN]-[LIVMF]-[DN]-x-E.

NAME: Chitinases family 19 signature 1.
CONSENSUS: C-x(4,5)-F-Y-[ST]-x(3)-[FY]-[LIVMF]-x-A-x(3)-[YF]-x(2)-F-[GSA].

NAME: Chitinases family 19 signature 2.
CONSENSUS: [LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W-[LIVM].

NAME: Alpha-lactalbumin / lysozyme C signature.
CONSENSUS: C-x(3)-C-x(2)-[LMF]-x(3)-[DEN]-[LI]-x(5)-C.

NAME: Alpha-galactosidase signature.
CONSENSUS: G-[LIVMFY]-x(2)-[LIVMFY]-x-[LIVM]-D-D-x-W-x(3,4)-R-[DNSF].

NAME: Trehalase signature 1.

CONSENSUS: P-G-G-R-F-x-E-x-Y-x-W-D-x-Y.

NAME: Trehalase signature 2.
CONSENSUS: Q-W-D-x-P-x-[GA]-W-[PA]-P.

NAME: Alpha-L-fucosidase putative active site.
CONSENSUS: P-x(2)-L-x(3)-K-W-E-x-C.

NAME: Glycosyl hydrolases family 1 active site.
CONSENSUS: [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN].

NAME: Glycosyl hydrolases family 1 N-terminal signature.
CONSENSUS: F-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x-[GSTA].

NAME: Glycosyl hydrolases family 2 signature 1.
CONSENSUS: N-x-[LIVMFYWD]-R-[STACN](2)-H-Y-P-x(4)-[LIVMFYW](2)-x(3)-[DN]-x(2)-
CONSENSUS: G-[LIVMFYW](4).

NAME: Glycosyl hydrolases family 2 acid/base catalyst.
CONSENSUS: [DENQF]-[KRVW]-N-H-[AP]-[SAC]-[LIVMF](3)-W-[GS]-x(2,3)-N-E.

NAME: Glycosyl hydrolases family 3 active site.
CONSENSUS: [LIVM](2)-[KR]-x-[EQK]-x(4)-G-[LIVMFT]-[LIVT]-[LIVMF]-[ST]-D-x(2)-
CONSENSUS: [SGADNI].

NAME: Glycosyl hydrolases family 5 signature.
CONSENSUS: [LIV]-[LIVMFYWGA](2)-[DNEQG]-[LIVMGST]-x-N-E-[PV]-[RHDNSTLIVFY].

NAME: Glycosyl hydrolases family 6 signature 1.
CONSENSUS: V-x-Y-x(2)-P-x-R-D-C-[GSAF]-x(2)-[GSA](2)-x-G.

NAME: Glycosyl hydrolases family 6 signature 2.
CONSENSUS: [LIVMYA]-[LIVA]-[LIVT]-[LIV]-E-P-D-[SAL]-[LI]-[PSAG].

NAME: Glycosyl hydrolases family 8 signature.
CONSENSUS: A-[ST]-D-[AG]-D-x(2)-[IM]-A-x-[SA]-[LIVM]-[LIVMG]-x-A-x(3)-[FW].

NAME: Glycosyl hydrolases family 9 active sites signature 1.
CONSENSUS: [STV]-x-[LIVMFY]-[STV]-x(2)-G-x-[NKR]-x(4)-[PLIVM]-H-x-R.

NAME: Glycosyl hydrolases family 9 active sites signature 2.
CONSENSUS: [FYW]-x-D-x(4)-[FYW]-x(3)-E-x-[STA]-x(3)-N-[STA].

NAME: Glycosyl hydrolases family 10 active site.
CONSENSUS: [GTA]-x(2)-[LIVN]-x-[IVMF]-[ST]-E-[LIY]-[DN]-[LIVMF].

NAME: Glycosyl hydrolases family 11 active site signature 1.
CONSENSUS: [PSA]-[LQ]-x-E-Y-Y-[LIVM](2)-[DE]-x-[FYWHN].

NAME: Glycosyl hydrolases family 11 active site signature 2.
CONSENSUS: [LIVMF]-x(2)-E-[AG]-[YWG]-[QRFGS]-[SG]-[STAN]-G-x-[SAF].

NAME: Glycosyl hydrolases family 16 active sites.
CONSENSUS: E-[LIV]-D-[LIV]-x(0,1)-E-x(2)-[GQ]-[KRNF]-x-[PSTA].

NAME: Glycosyl hydrolases family 17 signature.
CONSENSUS: [LIVM]-x-[LIVMFYWA](3)-[STAG]-E-[STA]-G-W-P-[STN]-x-[SAGQ].

NAME: Glycosyl hydrolases family 25 active sites signature.
CONSENSUS: D-[LIVM]-x(3)-[NQ]-[PG]-x(9,10)-G-x(4)-[LIVMFY](2)-K-x-[ST]-E-[GS]-x(2)-
CONSENSUS: Y-x-[DN].

NAME: Glycosyl hydrolases family 31 active site.
CONSENSUS: [GF]-[LIVMF]-W-x-D-M-[NSA]-E.

NAME: Glycosyl hydrolases family 31 signature 2.
CONSENSUS: G-[AV]-D-[LIVMT]-C-G-[FY]-x(3)-[ST]-x(3)-L-C-x-R-W-x(2)-[LV]-[GS]-[SA]-
CONSENSUS: F-x-P-F-x-R-[DN].

NAME: Glycosyl hydrolases family 32 active site.
CONSENSUS: H-x(2)-P-x(4)-[LIVM]-N-D-P-N-G.

NAME: Glycosyl hydrolases family 35 putative active site.
CONSENSUS: G-G-P-[LIVM](2)-x(2)-Q-x-E-N-E-[FY].

NAME: Glycosyl hydrolases family 39 active site.
CONSENSUS: W-x-F-E-x-W-N-E-P-[DN].

NAME: Glycosyl hydrolases family 45 active site.
CONSENSUS: [STA]-T-R-Y-[FYW]-D-x(5)-[CA].

NAME: Prokaryotic transglycosylases signature.
CONSENSUS: [LIVM]-x(3)-E-S-x(3)-[AP]-x(3)-S-x(5)-G-[LIVM]-[LIVMFYW]-x-[LIVMFYW]-
CONSENSUS: x(4)-[SAG].

NAME: Inosine-uridine preferring nucleoside hydrolase family signature.
CONSENSUS: D-x-D-[PT]-[GA]-x-D-D-[TAV]-[VI]-A.

NAME: Alkylbase DNA glycosidases alkA family signature.
CONSENSUS: G-I-G-x-W-[ST]-[AV]-x-[LIVMFY](2)-x-[LIVM]-x(8)-[MF]-x(2)-[ED]-D.

NAME: Formamidopyrimidine-DNA glycosylase signature.
CONSENSUS: C-x(2,4)-C-x-[GTAQ]-x-[IV]-x(7)-R-[GSTAN]-[STA]-x-[FYI]-C-x(2)-C-Q.

NAME: Uracil-DNA glycosylase signature.
CONSENSUS: [KR]-[LIV]-[LIVC]-[LIVM]-x-G-[QI]-D-P-Y.

NAME: S-adenosyl-L-homocysteine hydrolase signature 1.
CONSENSUS: [CS]-N-x-[FYL]-S-[ST]-[QA]-[DEN]-x-[AV](2)-A-A-[LIV]-[SAV].

NAME: S-adenosyl-L-homocysteine hydrolase signature 2.
CONSENSUS: G-K-x(3)-[LIV]-x-G-Y-G-x-V-G-[KR]-G-x-A.

NAME: Cytosol aminopeptidase signature.
CONSENSUS: N-T-D-A-E-G-R-L.

NAME: Aminopeptidase P and proline dipeptidase signature.
CONSENSUS: [HA]-[GSYR]-[LIVMT]-[SG]-H-x-[LIV]-G-[LIVM]-x-[IV]-H-[DE].

NAME: Methionine aminopeptidase subfamily 1 signature.
CONSENSUS: [MFY]-x-G-H-G-[LIVMC]-[GSH]-x(3)-H-x(4)-[LIVM]-x-[HN]-[YWV].

NAME: Methionine aminopeptidase subfamily 2 signature.
CONSENSUS: [DA]-[LIVMY]-x-K-[LIVM]-D-x-G-x-[HQ]-[LIVM]-[DNS]-G-x(3)-[DN].

NAME: Renal dipeptidase active site.
CONSENSUS: [LIVM]-E-G-[GA]-x(2)-[LIVMF]-x(6)-L-x(3)-Y-x(2)-G-[LIVM]-R.

NAME: Serine carboxypeptidases, serine active site.
CONSENSUS: [LIVM]-x-[GTA]-E-S-Y-[AG]-[GS].

NAME: Serine carboxypeptidases, histidine active site.
CONSENSUS: [LIVF]-x(2)-[LIVSTA]-x-[IVPST]-x-[GSDNQL]-[SAGV]-[SG]-H-x-[IVAQ]-P-x(3)-
CONSENSUS: [PSA].

NAME: Zinc carboxypeptidases, zinc-binding region 1 signature.
CONSENSUS: [PK]-x-[LIVMFY]-x-[LIVMFY]-x(4)-H-[STAG]-x-E-x-[LIVM]-[STAG]-x(6)-
CONSENSUS: [LIVMFYTA].

NAME: Zinc carboxypeptidases, zinc-binding region 2 signature.
CONSENSUS: H-[STAG]-x(3)-[LIVME]-x(2)-[LIVMFYW]-P-[FYW].

NAME: Serine proteases, trypsin family, histidine active site.
CONSENSUS: [LIVM]-[ST]-A-[STAG]-H-C.

NAME: Serine proteases, trypsin family, serine active site.
CONSENSUS: [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-[LIVMFYWH]-
CONSENSUS: [LIVMFYSTANQH].

NAME: Serine proteases, subtilase family, aspartic acid active site.
CONSENSUS: [STAIV]-x-[LIVMF]-[LIVM]-D-[DSTA]-G-[LIVMFC]-x(2,3)-[DNH].

NAME: Serine proteases, subtilase family, histidine active site.
CONSENSUS: H-G-[STM]-x-[VIC]-[STAGC]-[GS]-x-[LIVMA]-[STAGCLV]-[SAGM].

NAME: Serine proteases, subtilase family, serine active site.
CONSENSUS: G-T-S-x-[SA]-x-P-x(2)-[STAVC]-[AG].

NAME: Serine proteases, V8 family, histidine active site.
CONSENSUS: [ST]-G-[LIVMFYW](3)-[GN]-x(2)-T-[LIVM]-x-T-x(2)-H.

NAME: Serine proteases, V8 family, serine active site.
CONSENSUS: T-x(2)-[GC]-[NQ]-S-G-S-x-[LIVM]-[FY].

NAME: Serine proteases, omptin family signature 1.
CONSENSUS: W-T-D-x-S-x-H-P-x-T.

NAME: Serine proteases, omptin family signature 2.
CONSENSUS: A-G-Y-Q-E-[ST]-R-[FYW]-S-[FYW]-[TN]-A-x-G-G-[ST]-Y.

NAME: Prolyl endopeptidase family serine active site.
CONSENSUS: D-x(3)-A-x(3)-[LIVMFYW]-x(14)-G-x-S-x-G-G-[LIVMFYW](2).

NAME: Endopeptidase Clp serine active site.
CONSENSUS: T-x(2)-[LIVMF]-G-x-A-[SAC]-S-[MSA]-[PAG]-[STA].

NAME: Endopeptidase Clp histidine active site.
CONSENSUS: R-x(3)-[EAP]-x(3)-[LIVMFYT]-M-[LIVM]-H-Q-P.

NAME: ATP-dependent serine proteases, lon family, serine active site.
CONSENSUS: D-G-[PD]-S-A-[GS]-[LIVMCA]-[TA]-[LIVM].

NAME: Eukaryotic thiol (cysteine) proteases cysteine active site.
CONSENSUS: Q-x(3)-[GE]-x-C-[YW]-x(2)-[STAGC]-[STAGCV].

NAME: Eukaryotic thiol (cysteine) proteases histidine active site.
CONSENSUS: [LIVMGSTAN]-x-H-[GSACE]-[LIVM]-x-[LIVMAT](2)-G-x-[GSADNH].

NAME: Eukaryotic thiol (cysteine) proteases asparagine active site.
CONSENSUS: [FYCH]-[WI]-[LIVT]-x-[KRQAG]-N-[ST]-W-x(3)-[FYW]-G-x(2)-G-[LFYW]-
CONSENSUS: [LIVMFYG]-x-[LIVMF].

NAME: Ubiquitin carboxyl-terminal hydrolase family 1 cysteine active-site.
CONSENSUS: Q-x(3)-N-[SA]-C-G-x(3)-[LIVM](2)-H-[SA]-[LIVM]-[SA].

NAME: Ubiquitin carboxyl-terminal hydrolases family 2 signature 1.
CONSENSUS: G-[LIVMFY]-x(1,3)-[AGC]-[NASM]-x-C-[FYW]-[LIVMC]-[NST]-[SACV]-x-[LIVMS]-
CONSENSUS: Q.

NAME: Ubiquitin carboxyl-terminal hydrolases family 2 signature 2.
CONSENSUS: Y-x-L-x-[SAG]-[LIVMFT]-x(2)-H-x-G-x(4,5)-G-H-Y.

NAME: Caspase family histidine active site.
CONSENSUS: H-x(2,4)-[SC]-x(4)-[LIVMF](2)-[ST]-H-G.

NAME: Caspase family cysteine active site.
CONSENSUS: K-P-K-[LIVMF](4)-Q-A-C-[RQG]-G.

NAME: Eukaryotic and viral aspartyl proteases active site.
CONSENSUS: [LIVMFGAC]-[LIVMTADN]-[LIVFSA]-D-[ST]-G-[STAV]-[STAPDENQ]-x-[LIVMFSTNC]-
CONSENSUS: x-[LIVMFGTA].

NAME: Neutral zinc metallopeptidases, zinc-binding region signature.
CONSENSUS: [GSTALIVN]-x(2)-H-E-[LIVMFYW]-{DEHRKP}-H-x-[LIVMFYWGSPQ].

NAME: Matrixins cysteine switch.
CONSENSUS: P-R-C-[GN]-x-P-[DR]-[LIVSAPKQ].

NAME: Insulinase family, zinc-binding region signature.
CONSENSUS: G-x(8,9)-G-x-[STA]-H-[LIVMFY]-[LIVMC]-[DERN]-[HRKL]-[LMFAT]-x-[LFSTH]-x-
CONSENSUS: [GSTAN]-[GST].

//

AC PS01016;
DE Glycoprotease family signature.
CONSENSUS: [KR]-[GSAT]-x(4)-[FYWHL]-[DQNGK]-x-P-x-[LIVMFY]-x(3)-H-x(2)-[AG]-H-
CONSENSUS: [LIVM].

NAME: Proteasome A-type subunits signature.
CONSENSUS: [FY]-x(4)-[STNV]-x-[FYW]-S-P-x-G-[RKH]-x(2)-Q-[LIVM]-[DE]-Y-[SAD]-x(2)-
CONSENSUS: [SAG].

NAME: Proteasome B-type subunits signature.
CONSENSUS: [LIVMA]-[GSA]-[LIVMF]-x-[FYLVGAC]-x(2)-[GSACFY]-[LIVMSTAC](3)-[GAC]-
CONSENSUS: [GSTACV]-[DES]-x(15)-[RK]-x(12,13)-G-x(2)-[GSTA]-D.

NAME: Signal peptidases I serine active site.
CONSENSUS: [GS]-x-S-M-x-[PS]-[AT]-[LF].

NAME: Signal peptidases I lysine active site.
CONSENSUS: K-R-[LIVMSTA](2)-G-x-[PG]-G-[DE]-x-[LIVM]-x-[LIVMFY].

NAME: Signal peptidases I signature 3.
CONSENSUS: [LIVMFYW](2)-x(2)-G-D-[NH]-x(3)-[SND]-x(2)-[SG].

NAME: Signal peptidases II signature.
CONSENSUS: [GAF]-[GA]-[GAS]-[LIVM]-[GAS]-N-[LVMFG]-[LIVMFY]-D-R-[LIMFA].

NAME: Peptidase family U32 signature.
CONSENSUS: E-x-F-x(2)-G-[SA]-[LIVM]-C-x(4)-G-x-C-x-[LIVM]-S.

NAME: Amidases signature.
CONSENSUS: G-[GA]-S-S-[GS]-G-x-[GSA]-[GSAVY]-x-[LIVM]-[GSA]-x(6)-[GSA]-x-[GA]-x-D-
CONSENSUS: x-[GA]-x-S-[LIVM]-R-x-P-[GSAC].

NAME: Asparaginase / glutaminase active site signature 1.
CONSENSUS: [LIVM]-x(2)-T-G-G-T-[IV]-[AGS].

NAME: Asparaginase / glutaminase active site signature 2.
CONSENSUS: G-x-[LIVM]-x(2)-H-G-T-D-T-[LIVM].

NAME: Urease nickel ligands signature.
CONSENSUS: T-[AY]-[GA]-[GAT]-[LIVM]-D-x-H-[LIVM]-H-x(3)-P.

NAME: Urease active site.
CONSENSUS: [LIVM](2)-[CT]-H-[HN]-L-x(3)-[LIVM]-x(2)-D-[LIVM]-x-F-A.

NAME: ArgE / dapE / ACY1 / CPG2 / yscS family signature 1.
CONSENSUS: [LIV]-[GALMY]-[LIVMF]-x-[GSA]-H-x-D-[TV]-[STAV].

NAME: ArgE / dapE / ACY1 / CPG2 / yscS family signature 2.
CONSENSUS: [GSTAI]-[SANQ]-D-x-K-[GSACN]-x(2)-[LIVMA]-x(2)-[LIVMFY]-x(14,17)-[LIVM]-
CONSENSUS: x-[LIVMF]-[LIVMSTAG]-[LIVMFA]-x(2)-[DNG]-E-E-x-[GSTN].

NAME: Dihydroorotase signature 1.
CONSENSUS: D-[LIVMFYWSAP]-H-[LIVA]-H-[LIVF]-[RN]-x-[PGN].

NAME: Dihydroorotase signature 2.
CONSENSUS: [GA]-[ST]-D-x-A-P-H-x(4)-K.

NAME: Beta-lactamase class-A active site.
CONSENSUS: [FY]-x-[LIVMFY]-x-S-[TV]-x-K-x(4)-[AGLM]-x(2)-[LC].

NAME: Beta-lactamase class-C active site.
CONSENSUS: F-E-[LIVM]-G-S-[LIVMG]-[SA]-K.

NAME: Beta-lactamase class-D active site.
CONSENSUS: [PA]-x-S-[ST]-F-K-[LIV]-[PAL]-x-[STA]-[LI].

NAME: Beta-lactamases class B signature 1.
CONSENSUS: [LI]-x-[STN]-[HN]-x-H-[GSTA]-D-x(2)-G-[GP]-x(7,8)-[GS].

NAME: Beta-lactamases class B signature 2.
CONSENSUS: P-x(3)-[LIVM](2)-x-G-x-C-[LIVMF](2)-K.

NAME: Arginase family signature 1.
CONSENSUS: [LIVMF]-G-G-x-H-x-[LIVMT]-[STAV]-x-[PAG]-x(3)-[GSTA].

NAME: Arginase family signature 2.
CONSENSUS: [LIVM](2)-x-[LIVMFY]-D-[AS]-H-x-D.

NAME: Arginase family signature 3.
CONSENSUS: [ST]-[LIVMFY]-D-[LIVM]-D-x(3)-[PAQ]-x(3)-P-[GSA]-x(7)-G.

NAME: Adenosine and AMP deaminase signature.

CONSENSUS: [SA]-[LIVM]-[NGS]-[STA]-D-D-P.

NAME: Cytidine and deoxycytidylate deaminases zinc-binding region signature.
CONSENSUS: [CH]-[AGV]-E-x(2)-[LIVMFGAT]-[LIVM]-x(17,33)-P-C-x(2,8)-C-x(3)-[LIVM].

NAME: GTP cyclohydrolase I signature 1.
CONSENSUS: [EN]-[LIVM](2)-x(2)-[KRQN]-[DN]-[LIVM]-x(3)-[ST]-x-C-E-H-H.

NAME: GTP cyclohydrolase I signature 2.
CONSENSUS: [SA]-x-[RK]-x-Q-[LIVM]-Q-E-[RN]-[LI]-[TSN].

NAME: Nitrilases / cyanide hydratase signature 1.
CONSENSUS: G-x(2)-[LIVMFY](2)-x-[IF]-x-E-x(2)-[LIVM]-x-G-Y-P.

NAME: Nitrilases / cyanide hydratase active site signature.
CONSENSUS: G-[GAQ]-x(2)-C-[WA]-E-[NH]-x(2)-[PST]-[LIVMFYS]-x-[KR].

NAME: Inorganic pyrophosphatase signature.
CONSENSUS: D-[SGDN]-D-[PE]-[LIVMF]-D-[LIVMGAC].

NAME: Acylphosphatase signature 1.
CONSENSUS: [LIV]-x-G-x-V-Q-G-V-x-[FM]-R.

NAME: Acylphosphatase signature 2.
CONSENSUS: G-[FYW]-[AVC]-[KRQAM]-N-x(3)-G-x-V-x(5)-G.

NAME: ATP synthase alpha and beta subunits signature.
CONSENSUS: P-[SAP]-[LIV]-[DNH]-x(3)-S-x-S.

NAME: ATP synthase gamma subunit signature.
CONSENSUS: [IV]-T-x-E-x(2)-[DE]-x(3)-G-A-x-[SAKR].

NAME: ATP synthase delta (OSCP) subunit signature.
CONSENSUS: [LIVM]-x-[LIVMFYT]-x(3)-[LIVMT]-[DENQK]-x(2)-[LIVM]-x-[GSA]-G-[LIVMFYGA]-
CONSENSUS: x-[LIVM]-[KRHENQ]-x-[GSEN].

NAME: ATP synthase a subunit signature.
CONSENSUS: [STAGN]-x-[STAG]-[LIVMF]-R-L-x-[SAGV]-N-[LIVMT].

NAME: ATP synthase c subunit signature.
CONSENSUS: [GSTA]-R-[NQ]-P-x(10)-[LIVMFYW](2)-x(3)-[LIVMFYW]-x-[DE].

NAME: E1-E2 ATPases phosphorylation site.
CONSENSUS: D-K-T-G-T-[LI]-[TI].

NAME: Sodium and potassium ATPases beta subunits signature 1.
CONSENSUS: [FYW]-x(2)-[FYW]-x-[FYW]-[DN]-x(6)-[LIVM]-G-R-T-x(3)-W.

NAME: Sodium and potassium ATPases beta subunits signature 2.
CONSENSUS: [RK]-x(2)-C-[RKQWI]-x(5)-L-x(2)-C-[SA]-G.

NAME: GDA1/CD39 family of nucleoside phosphatases signature.
CONSENSUS: [LIVM]-x-G-x(2)-E-G-x-[FY]-x-[FW]-[LIVA]-[TAG]-x-N-[HY].

NAME: Iodothyronine deiodinases active site.
CONSENSUS: R-P-L-V-x-N-F-G-S-[CA]-T-C-P-x-F.

NAME: Cutinase, serine active site.
CONSENSUS: P-x-[STA]-x-[LIV]-[IVT]-x-[GS]-G-Y-S-[QL]-G.

NAME: Cutinase, aspartate and histidine active sites.
CONSENSUS: C-x(3)-D-x-[IV]-C-x-G-[GST]-x(2)-[LIVM]-x(2,3)-H.

NAME: DDC / GAD / HDC / TyrDC pyridoxal-phosphate attachment site.
CONSENSUS: S-[LIVMFYW]-x(5)-K-[LIVMFYWG](2)-x(3)-[LIVMFYW]-x-[CA]-x(2)-[LIVMFYWQ]-
CONSENSUS: x(2)-[RK].

NAME: Orn/Lys/Arg decarboxylases family 1 pyridoxal-P attachment site.
CONSENSUS: [STAV]-x-S-x-H-K-x(2)-[GSTAN](2)-x-[STA]-Q-[STA](2).

NAME: Orn/DAP/Arg decarboxylases family 2 pyridoxal-P attachment site.
CONSENSUS: [FY]-[PA]-x-K-[SACV]-[NHCLFW]-x(4)-[LIVMF]-[LIVMTA]-x(2)-[LIVMA]-x(3)-
CONSENSUS: [GTE].

NAME: Orn/DAP/Arg decarboxylases family 2 signature 2.
CONSENSUS: [GS]-x(2,6)-[LIVMSCP]-x(2)-[LIVMF]-[DNS]-[LIVMCA]-G-G-G-[LIVMFY]-
CONSENSUS: [GSTPCEQ].

NAME: Orotidine 5'-phosphate decarboxylase active site.
CONSENSUS: [LIVMFTA]-[LIVMF]-x-D-x-K-x(2)-D-I-[GP]-x-T-[LIVMTA].

NAME: Phosphoenolpyruvate carboxylase active site 1.
CONSENSUS: [VT]-x-T-A-H-P-T-[EQ]-x(2)-R-[KRH].

NAME: Phosphoenolpyruvate carboxylase active site 2.
CONSENSUS: [IV]-M-[LIVM]-G-Y-S-D-S-x-K-D-[STAG]-G.

NAME: Phosphoenolpyruvate carboxykinase (GTP) signature.
CONSENSUS: F-P-S-A-C-G-K-T-N.

NAME: Phosphoenolpyruvate carboxykinase (ATP) signature.
CONSENSUS: L-I-G-D-D-E-H-x-W-x-[DE]-x-G-[IV]-x-N.

NAME: Uroporphyrinogen decarboxylase signature 1.
CONSENSUS: P-x-W-x-M-R-Q-A-G-R.

NAME: Uroporphyrinogen decarboxylase signature 2.
CONSENSUS: G-F-[STAGCV]-[STAGC]-x-P-[FYW]-T-[LV]-x(2)-Y-x(2)-[AE]-[GK].

NAME: Indole-3-glycerol phosphate synthase signature.
CONSENSUS: [LIVMFY]-[LIVMC]-x-E-[LIVMFYC]-K-[KRSP]-[STAK]-S-P-[ST]-x(3)-[LIVMFYST].

NAME: Ribulose bisphosphate carboxylase large chain active site.
CONSENSUS: G-x-[DN]-F-x-K-x-D-E.

NAME: Fructose-bisphosphate aldolase class-I active site.
CONSENSUS: [LIVM]-x-[LIVMFYW]-E-G-x-[LS]-L-K-P-[SN].

NAME: Fructose-bisphosphate aldolase class-II signature 1.
CONSENSUS: [FYVM]-x(1,3)-[LIVMH]-[APN]-[LIVM]-x(1,2)-[LIVM]-H-x-D-H-[GACH].

NAME: Fructose-bisphosphate aldolase class-II signature 2.
CONSENSUS: [LIVM]-E-x-E-[LIVM]-G-x(2)-[GM]-[GSTA]-x-E.

NAME: Malate synthase signature.
CONSENSUS: [KR]-[DENQ]-H-x(2)-G-L-N-x-G-x-W-D-Y-[LIVM]-F.

NAME: Hydroxymethylglutaryl-coenzyme A lyase active site.
CONSENSUS: S-V-A-G-L-G-G-C-P-Y.

NAME: Hydroxymethylglutaryl-coenzyme A synthase active site.
CONSENSUS: N-x-[DN]-[IV]-E-G-[IV]-D-x(2)-N-A-C-[FY]-x-G.

NAME: Citrate synthase signature.
CONSENSUS: G-[FYA]-[GA]-H-x-[IV]-x(1,2)-[RKT]-x(2)-D-[PS]-R.

NAME: Alpha-isopropylmalate and homocitrate synthases signature 1.
CONSENSUS: L-R-[DE]-G-x-Q-x(10)-K.

NAME: Alpha-isopropylmalate and homocitrate synthases signature 2.
CONSENSUS: [LIVMFW]-x(2)-H-x-H-[DN]-D-x-G-x-[GAS]-x-[GASLI].

NAME: KDPG and KHG aldolases active site.
CONSENSUS: G-[LIVM]-x(3)-E-[LIV]-T-[LF]-R.

NAME: KDPG and KHG aldolases Schiff-base forming residue.
CONSENSUS: G-x(3)-[LIVMF]-K-[LF]-F-P-[SA]-x(3)-G.

NAME: Isocitrate lyase signature.
CONSENSUS: K-[KR]-C-G-H-[LMQ].

NAME: Beta-eliminating lyases pyridoxal-phosphate attachment site.
CONSENSUS: Y-x-D-x(3)-M-S-[GA]-K-K-D-x-[LIVM](2)-x-[LIVM]-G-G.

NAME: DNA photolyases class 1 signature 1.
CONSENSUS: T-G-x-P-[LIVM](2)-D-A-x-M-[RA]-x-[LIVM].

NAME: DNA photolyases class 1 signature 2.

CONSENSUS: [DN]-R-x-R-[LIVM](2)-x-[STA](2)-F-[LIVMFA]-x-K-x-L-x(2,3)-W-[KRQ].

NAME: DNA photolyases class 2 signature 1.
CONSENSUS: F-x-E-E-x-[LIVM](2)-R-R-E-L-x(2)-N-F.

NAME: DNA photolyases class 2 signature 2.
CONSENSUS: G-x-H-D-x(2)-W-x-E-R-x-[LIVM]-F-G-K-[LIVM]-R-[FY]-M-N.

NAME: Eukaryotic-type carbonic anhydrases signature.
CONSENSUS: S-E-H-x-[LIVM]-x(4)-[FYH]-x(2)-E-[LIVM]-H-[LIVMFA](2).

NAME: Prokaryotic-type carbonic anhydrases signature 1.
CONSENSUS: C-[SA]-D-S-R-[LIVM]-x-[AP].

NAME: Prokaryotic-type carbonic anhydrases signature 2.
CONSENSUS: [EQ]-Y-A-[LIVM]-x(2)-[LIVM]-x(4)-[LIVMF](3)-x-G-H-x(2)-C-G.

NAME: Fumarate lyases signature.
CONSENSUS: G-S-x(2)-M-x(2)-K-x-N.

NAME: Aconitase family signature 1.
CONSENSUS: [LIVM]-x(2)-[GSACIVM]-x-[LIV]-[GTIV]-[STP]-C-x(0,1)-T-N-[GSTANI]-x(4)-
CONSENSUS: [LIVMA].

NAME: Aconitase family signature 2.
CONSENSUS: G-x(2)-[LIVWPQ]-x(3)-[GAC]-C-[GSTAM]-[LIMPTA]-C-[LIMV]-[GA].

NAME: Dihydroxy-acid and 6-phosphogluconate dehydratases signature 1.
CONSENSUS: C-D-K-x(2)-P-[GA]-x(3)-[GA].

NAME: Dihydroxy-acid and 6-phosphogluconate dehydratases signature 2.
CONSENSUS: [SA]-L-[LIVM]-T-D-[GA]-R-[LIVMF]-S-[GA]-[GAV]-[ST].

NAME: Dehydroquinase class I active site.
CONSENSUS: D-[LIVM]-[DE]-[LIVN]-x(18,20)-[LIVM](2)-x-[SC]-[NHY]-H-[DN].

NAME: Dehydroquinase class II signature.
CONSENSUS: [LIVM]-[NQ]-G-P-N-[LV]-x(2)-L-G-x-R-[QED]-P-x(2)-[FY]-G.

NAME: Enolase signature.
CONSENSUS: [LIV](3)-K-x-N-Q-I-G-[ST]-[LIV]-[ST]-[DE]-[STA].

NAME: Serine/threonine dehydratases pyridoxal-phosphate attachment site.
CONSENSUS: [DESH]-x(4,5)-[STVG]-x-[AS]-[FYI]-K-[DLIFSA]-[RVMF]-[GA]-[LIVMGA].

NAME: Enoyl-CoA hydratase/isomerase signature.
CONSENSUS: [LIVM]-[STA]-x-[LIVM]-[DENQRHSTA]-G-x(3)-[AG](3)-x(4)-[LIVMST]-x-[CSTA]-
CONSENSUS: [DQHP]-[LIVMFY].

NAME: Imidazoleglycerol-phosphate dehydratase signature 1.
CONSENSUS: [LIVMY]-[DE]-x-H-H-x(2)-E-x(2)-[GCA]-[LIVM]-[STAC]-[LIVM].

NAME: Imidazoleglycerol-phosphate dehydratase signature 2.
CONSENSUS: G-x-[DN]-x-H-H-x(2)-E-[STAGC]-x-[FY]-K.

NAME: Tryptophan synthase alpha chain signature.
CONSENSUS: [LIVM]-E-[LIVM]-G-x(2)-[FYC]-[ST]-[DE]-[PA]-[LIVMY]-[AGLI]-[DE]-G.

NAME: Tryptophan synthase beta chain pyridoxal-phosphate attachment site.
CONSENSUS: [LIVM]-x-H-x-G-[STA]-H-K-x-N.

NAME: Delta-aminolevulinic acid dehydratase active site.
CONSENSUS: G-x-D-x-[LIVM](2)-[IV]-K-P-[GSA]-x(2)-Y.

NAME: Urocanase active site.
CONSENSUS: F-Q-G-L-P-x-R-I-C-W.

NAME: Prephenate dehydratase signature 1.
CONSENSUS: [FY]-x-[LIVM]-x(2)-[LIVM]-x(5)-[DN]-x(5)-T-R-F-[LIVMW]-x-[LIVM].

NAME: Prephenate dehydratase signature 2.
CONSENSUS: [LIVM]-[ST]-[KR]-[LIVM]-E-[ST]-R-P.

NAME: Dihydrodipicolinate synthetase signature 1.

CONSENSUS: [GSA]-[LIVM]-[LIVMFY]-x(2)-G-[ST]-[TG]-G-E-[GASNF]-x(6)-[EQ].

NAME: Dihydrodipicolinate synthetase signature 2.
CONSENSUS: Y-[DNS]-[LIVMF]-P-x(2)-[ST]-x(3)-[LIVM]-x(13,14)-[LIVM]-x-[SGA]-[LIVMF]-
CONSENSUS: K-[DEQAF]-[STAC].

NAME: RsuA family of pseudouridine synthase signature.
CONSENSUS: G-R-L-D-x(2)-[ST]-x-G-[LIVMF](4)-[ST]-[DNT].

NAME: Cysteine synthase/cystathionine beta-synthase P-phosphate attachment site.
CONSENSUS: K-x-E-x(3)-[PA]-[STAGC]-x-S-[IVAP]-K-x-R-x-[STAG]-x(2)-[LIVM].

NAME: Phenylalanine and histidine ammonia-lyases signature.
CONSENSUS: G-[STG]-[LIVM]-[STG]-[AC]-S-G-[DH]-L-x-P-L-[SA]-x(2)-[SA].

NAME: Porphobilinogen deaminase cofactor-binding site.
CONSENSUS: E-R-x-[LIVMFA]-x(3)-[LIVMF]-x-G-[GSA]-C-x-[IVT]-P-[LIVMF]-[GSA].

NAME: Cys/Met metabolism enzymes pyridoxal-phosphate attachment site.
CONSENSUS: [DQ]-[LIVMF]-x(3)-[STAGC]-[STAGCI]-T-K-[FYWQ]-[LIVMF]-x-G-[HQ]-[SGNH].

NAME: Glyoxalase I signature 1.
CONSENSUS: [HQ]-[IVT]-x-[LIVFY]-x-[IV]-x(5)-[STA]-x(2)-F-[YM]-x(2,3)-[LMF]-G-[LMF].

NAME: Glyoxalase I signature 2.
CONSENSUS: G-[NTKQ]-x(0,5)-[GA]-[LVFY]-[GH]-H-[IVF]-[CGA]-x-[STAGL]-x(2)-[DNC].

NAME: Cytochrome c and c1 heme lyases signature 1.
CONSENSUS: H-N-x(2)-N-E-x(2)-W-[NQKR]-x(4)-W-E.

NAME: Cytochrome c and c1 heme lyases signature 2.
CONSENSUS: P-F-D-R-H-D-W.

NAME: Adenylate cyclases class-I signature 1.
CONSENSUS: E-Y-F-G-[SA](2)-L-W-x-L-Y-K.

NAME: Adenylate cyclases class-I signature 2.
CONSENSUS: Y-R-N-x-W-[NS]-E-[LIVM]-R-T-L-H-F-x-G.

NAME: Guanylate cyclases signature.
CONSENSUS: G-V-[LIVM]-x(0,1)-G-x(5)-[FY]-x-[LIVM]-[FYW]-[GS]-[DNTHKW]-[DNT]-[IV]-
CONSENSUS: [DNTA]-x(5)-[DE].

NAME: Chorismate synthase signature 1.
CONSENSUS: G-E-S-H-[GC]-x(2)-[LIVM]-[GTV]-x-[LIVM](2)-[DE]-G-x-[PV].

NAME: Chorismate synthase signature 2.
CONSENSUS: [GE]-R-[SA](2)-[SAG]-R-[EV]-[ST]-x(2)-[RH]-V-x(2)-G.

NAME: Chorismate synthase signature 3.
CONSENSUS: R-[SH]-D-[PSV]-[CSAV]-x(4)-[GAI]-x-[IVGSP]-[LIVM]-x-E-[STAH]-[LIVM].

NAME: 6-pyruvoyl tetrahydropterin synthase signature 1.
CONSENSUS: C-N-N-x(2)-G-H-G-H-N-Y.

NAME: 6-pyruvoyl tetrahydropterin synthase signature 2.
CONSENSUS: D-H-K-N-L-D-x-D.

NAME: Ferrochelatase signature.
CONSENSUS: [LIVMF](2)-x-S-x-H-[GS]-[LIVM]-P-x(4,5)-[DENQKR]-x-G-D-x-Y.

NAME: Alanine racemase pyridoxal-phosphate attachment site.
CONSENSUS: V-x-K-A-[DN]-[GA]-Y-G-H-G.

NAME: Aspartate and glutamate racemases signature 1.
CONSENSUS: [IVA]-[LIVM]-x-C-x(0,1)-N-[ST]-[MSA]-[STH]-[LIVFYSTANK].

NAME: Aspartate and glutamate racemases signature 2.
CONSENSUS: [LIVM](2)-x-[AG]-C-T-[DEH]-[LIVMFY]-[PNGRS]-x-[LIVM].

NAME: Mandelate racemase / muconate lactonizing enzyme family signature 1.
CONSENSUS: A-x-[SAG](2)-[LIVM]-[DE]-x-A-x(2)-D-x(2)-[GA]-[KR].

NAME: Mandelate racemase / muconate lactonizing enzyme family signature 2.

CONSENSUS: G-x(7)-D-x(9)-A-x(14)-[LIVM]-E-[DENQ]-P-x(4)-[DENQ].

NAME: Ribulose-phosphate 3-epimerase family signature 1.
CONSENSUS: [LIVMF]-H-[LIVMFY]-D-[LIVM]-x-D-x(1,2)-[FY]-[LIVM]-x-N-x-[STAV].

NAME: Ribulose-phosphate 3-epimerase family signature 2.
CONSENSUS: [LIVMA]-x-[LIVM]-M-[ST]-[VS]-x-P-x(3)-G-Q-x-F-x(6)-[NK]-[LIVMC].

NAME: Aldose 1-epimerase putative active site.
CONSENSUS: [NS]-x-T-N-H-x-Y-[FW]-N-[LI].

NAME: Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature.
CONSENSUS: [FY]-x(2)-[STCNLV]-x-F-H-[RH]-[LIVMN]-[LIVM]-x(2)-F-[LIVM]-x-Q-[AG]-G.

NAME: Cyclophilin-type peptidyl-prolyl cis-trans isomerase profile.

NAME: FKBP-type peptidyl-prolyl cis-trans isomerase signature 1.
CONSENSUS: [LIVMC]-x-[YF]-x-[GVL]-x(1,2)-[LFT]-x(2)-G-x(3)-[DE]-[STAEQK]-[STAN].

NAME: FKBP-type peptidyl-prolyl cis-trans isomerase signature 2.
CONSENSUS: [LIVMFY]-x(2)-[GA]-x(3,4)-[LIVMF]-x(2)-[LIVMFHK]-x(2)-G-x(4)-[LIVMF]-
CONSENSUS: x(3)-[PSGAQ]-x(2)-[AG]-[FY]-G.

NAME: FKBP-type peptidyl-prolyl cis-trans isomerase domain profile.

NAME: PpiC-type peptidyl-prolyl cis-trans isomerase signature.
CONSENSUS: F-[GSADEI]-x-[LVAQ]-A-x(3)-[ST]-x(3,4)-[STQ]-x(3,5)-[GER]-G-x-[LIVM]-
CONSENSUS: [GS].

NAME: Triosephosphate isomerase active site.
CONSENSUS: [AV]-Y-E-P-[LIVM]-W-[SA]-I-G-T-[GK].

NAME: Xylose isomerase signature 1.
CONSENSUS: [LI]-E-P-K-P-x(2)-P.

NAME: Xylose isomerase signature 2.
CONSENSUS: [FL]-H-D-x-D-[LIV]-x-[PD]-x-[GDE].

NAME: Phosphomannose isomerase type I signature 1.
CONSENSUS: Y-x-D-x-N-H-K-P-E.

NAME: Phosphomannose isomerase type I signature 2.
CONSENSUS: H-A-Y-[LIVM]-x-G-x(2)-[LIVM]-E-x-M-A-x-S-D-N-x-[LIVM]-R-A-G-x-T-P-K.

NAME: Phosphoglucose isomerase signature 1.
CONSENSUS: [DENS]-x-[LIVM]-G-G-R-[FY]-S-[LIVMT]-x-[STA]-[PSAC]-[LIVMA]-G.

NAME: Phosphoglucose isomerase signature 2.
CONSENSUS: [GS]-x-[LIVM]-[LIVMFYW]-x(4)-[FY]-[DN]-Q-x-G-V-E-x(2)-K.

NAME: Glucosamine/galactosamine-6-phosphate isomerases signature.
CONSENSUS: [LIVM]-x(3)-G-x-[LIT]-x-[LIV]-x-[LIVM]-x-G-[LIVM]-G-x-[DEN]-G-H.

NAME: Phosphoglycerate mutase family phosphohistidine signature.
CONSENSUS: [LIVM]-x-R-H-G-[EQ]-x(3)-N.

NAME: Phosphoglucomutase and phosphomannomutase phosphoserine signature.
CONSENSUS: [GSA]-[LIVM]-x-[LIVM]-[ST]-[PGA]-S-H-x-P-x(4)-[GNHE].

NAME: Methylmalonyl-CoA mutase signature.
CONSENSUS: R-I-A-R-N-[TQ]-x(2)-[LIVMFY](2)-x-[EQ]-E-x(4)-[KRN]-x(2)-D-P-x-[GSA]-
CONSENSUS: G-S.

NAME: Terpene synthases signature.
CONSENSUS: [DE]-G-S-W-x-G-x-W-[GA]-[LIVM]-x-[FY]-x-Y-[GA].

NAME: Eukaryotic DNA topoisomerase I active site.
CONSENSUS: [DEN]-x(6)-[GS]-[IT]-S-K-x(2)-Y-[LIVM]-x(3)-[LIVM].

NAME: Prokaryotic DNA topoisomerase I active site.
CONSENSUS: [EQ]-x-L-Y-[DEQT]-x(3,12)-[LI]-[ST]-Y-x-R-[ST]-[DEQS].

NAME: DNA topoisomerase II signature.
CONSENSUS: [LIVMA]-x-E-G-[DN]-S-A-x-[STAG].

NAME: Aminoacyl-transfer RNA synthetases class-I signature.
CONSENSUS: P-x(0,2)-[GSTAN]-[DENQGAPK]-x-[LIVMFP]-[HT]-[LIVMYAC]-G-[HNTG]-
CONSENSUS: [LIVMFYSTAGPC].

NAME: Aminoacyl-transfer RNA synthetases class-II signature 1.
CONSENSUS: [FYH]-R-x-[DE]-x(4,12)-[RH]-x(3)-F-x(3)-[DE].

NAME: Aminoacyl-transfer RNA synthetases class-II signature 2.
CONSENSUS: [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY].

NAME: WHEP-TRS domain signature.
CONSENSUS: [QY]-G-[DNEA]-x-[LIV]-[KR]-x(2)-K-x(2)-[KRNG]-[AS]-x(4)-[LIV]-[DENK]-
CONSENSUS: x(2)-[IV]-x(2)-L-x(3)-K.

NAME: ATP-citrate lyase / succinyl-CoA ligases family signature 1.
CONSENSUS: S-[KR]-S-G-[GT]-[LIVM]-[GST]-x-[EQ]-x(8,10)-G-x(4)-[LIVM]-[GA]-[LIVM]-G-
CONSENSUS: G-D.

NAME: ATP-citrate lyase / succinyl-CoA ligases family active site.
CONSENSUS: G-x(2)-A-x(4,7)-[RQT]-[LIVMF]-G-H-[AS]-[GH].

NAME: ATP-citrate lyase / succinyl-CoA ligases family signature 3.
CONSENSUS: G-x-[IV]-x(2)-[LIVMF]-x-[NA]-G-[GA]-G-[LA]-[STAV]-x(4)-D-x-[LIVM]-x(3)-
CONSENSUS: G-[GRE].

NAME: Glutamine synthetase signature 1.
CONSENSUS: [FYWL]-D-G-S-S-x(6,8)-[DENQSTAK]-[SA]-[DE]-x(2)-[LIVMFY].

NAME: Glutamine synthetase putative ATP-binding region signature.
CONSENSUS: K-P-[LIVMFYA]-x(3,5)-[NPAT]-G-[GSTAN]-G-x-H-x(3)-S.

NAME: Glutamine synthetase class-I adenylation site.
CONSENSUS: K-[LIVM]-x(5)-[LIVMA]-D-[RK]-[DN]-[LI]-Y.

NAME: D-alanine--D-alanine ligase signature 1.
CONSENSUS: H-G-x(2)-G-E-D-G-x-[LIVMA]-[QSA]-[GSA].

NAME: D-alanine--D-alanine ligase signature 2.
CONSENSUS: [LIV]-x(3)-[GA]-x-[GSAIV]-R-[LIVCA]-D-[LIVMF](2)-x(7,9)-[LI]-x-E-
CONSENSUS: [LIVA]-N-[STP]-x-P-[GA].

NAME: SAICAR synthetase signature 1.
CONSENSUS: [LIVMF](2)-P-[LIVM]-E-x-[LIVM]-[LIVMCA]-R-x(3)-[TA]-G-S.

NAME: SAICAR synthetase signature 2.
CONSENSUS: [LIVM]-[LIVMA]-D-x-K-[LIVMFY]-E-F-G.

NAME: Folylpolyglutamate synthase signature 1.
CONSENSUS: [LIVMFY]-x-[LIVM]-[STAG]-G-T-[NK]-G-K-x-[ST]-x(7)-[LIVM](2)-x(3)-[GSK].

NAME: Folylpolyglutamate synthase signature 2.
CONSENSUS: [LIVMFY](2)-E-x-G-[LIVM]-[GA]-G-x(2)-D-x-[GST]-x-[LIVM](2).

NAME: Ubiquitin-activating enzyme signature 1.
CONSENSUS: K-A-C-S-G-K-F-x-P.

NAME: Ubiquitin-activating enzyme active site.
CONSENSUS: P-[LIVM]-C-T-[LIVM]-[KRH]-x-[FT]-P.

NAME: Ubiquitin-conjugating enzymes active site.
CONSENSUS: [FYWLSP]-H-[PC]-[NH]-[LIV]-x(3,4)-G-x-[LIV]-C-[LIV]-x-[LIV].

NAME: Formate--tetrahydrofolate ligase signature 1.
CONSENSUS: G-[LIVM]-K-G-G-A-A-G-G-G-Y.

NAME: Formate--tetrahydrofolate ligase signature 2.
CONSENSUS: V-A-T-[IV]-R-A-L-K-x-[HN]-G-G.

NAME: Adenylosuccinate synthetase GTP-binding site.
CONSENSUS: Q-W-G-D-E-G-K-G.

NAME: Adenylosuccinate synthetase active site.
CONSENSUS: G-I-[GR]-P-x-Y-x(2)-K-x(2)-R.

NAME: Argininosuccinate synthase signature 1.
CONSENSUS: A-[FY]-S-G-G-L-D-T-S.

NAME: Argininosuccinate synthase signature 2.
CONSENSUS: G-x-T-x-K-G-N-D-x(2)-R-F.

NAME: Phosphoribosylglycinamide synthetase signature.
CONSENSUS: R-F-G-D-P-E-x-[QM].

NAME: Carbamoyl-phosphate synthase subdomain signature 1.
CONSENSUS: [FYV]-[PS]-[LIVMC]-[LIVMA]-[LIVM]-[KR]-[PSA]-[STA]-x(3)-[SG]-G-x-[AG].

NAME: Carbamoyl-phosphate synthase subdomain signature 2.
CONSENSUS: [LIVMF]-[LIMN]-E-[LIVMCA]-N-[PATLIVM]-[KR]-[LIVMSTAC].

NAME: ATP-dependent DNA ligase AMP-binding site.
CONSENSUS: [EDQH]-x-K-x-[DN]-G-x-R-[GACIVM].

NAME: ATP-dependent DNA ligase signature 2.
CONSENSUS: E-G-[LIVMA]-[LIVM](2)-[KR]-x(5,8)-[YW]-[QNEK]-x(2,6)-[KRH]-x(3,5)-K-
CONSENSUS: [LIVMFY]-K.

NAME: NAD-dependent DNA ligase signature 1.
CONSENSUS: K-[LIVM]-D-G-[LIVM]-[SA]-x(4)-Y-x(2)-G-x-L-x(4)-[ST]-R-G-[DN]-G-x(2)-G-
CONSENSUS: [DE]-[DENL].

NAME: NAD-dependent DNA ligase signature 2.
CONSENSUS: [IV]-G-[KR]-[ST]-G-x-[LIVM]-[STNK]-x-[VT]-x(2)-L-x-[PS]-V.

NAME: RNA 3'-terminal phosphate cyclase signature.
CONSENSUS: [RH]-G-x(2)-P-x-G(3)-x-[LIV].

NAME: Lipoate-protein ligase B signature.
CONSENSUS: R-G-G-x(2)-T-[FYW]-H-x(2)-[GH]-Q-x-[LIV]-x-Y.

NAME: Isopenicillin N synthetase signature 1.
CONSENSUS: [RK]-x-[STA]-x(2)-S-x-C-Y-[SL].

NAME: Isopenicillin N synthetase signature 2.
CONSENSUS: [LIVM](2)-x-C-G-[STA]-x(2)-[STAG]-x(2)-T-x-[DNG].

NAME: Site-specific recombinases active site.
CONSENSUS: Y-[LIVAC]-R-[VA]-S-[ST]-x(2)-Q.

NAME: Site-specific recombinases signature 2.
CONSENSUS: G-[DE]-x(2)-[LIVM]-x(3)-[LIVM]-[DT]-R-[LIVM]-[GSA].

NAME: Transposases, Mutator family, signature.
CONSENSUS: D-x(3)-G-[LIVMF]-x(6)-[STAV]-[LIVMFYW]-[PT]-x-[STAV]-x(2)-[QR]-x-C-x(2)-
CONSENSUS: H.

NAME: Transposases, IS30 family, signature.
CONSENSUS: R-G-x(2)-E-N-x-N-G-[LIVM](2)-R-[QE]-[LIVMFY](2)-P-K.

NAME: Autoinducers synthetases family signature.
CONSENSUS: [LMFY]-R-x(3)-F-x(2)-[KR]-x(2)-W-x-[LIVM]-x(6,9)-E-x-D-x-[FY]-D.

NAME: Thiamine pyrophosphate enzymes signature.
CONSENSUS: [LIVMF]-[GSA]-x(5)-P-x(4)-[LIVMFYW]-x-[LIVMF]-x-G-D-[GSA]-[GSAC].

NAME: Biotin-requiring enzymes attachment site.
CONSENSUS: [GN]-[DEQTR]-x-[LIVMFY]-x(2)-[LIVM]-x-[AIV]-M-K-[LMAT]-x(3)-[LIVM]-x-
CONSENSUS: [SAV].

NAME: 2-oxo acid dehydrogenases acyltransferase component lipoyl binding site.
CONSENSUS: [GN]-x(2)-[LIVF]-x(5)-[LIVFC]-x(2)-[LIVFA]-x(3)-K-[STAIV]-[STAVQDN]-
CONSENSUS: x(2)-[LIVMFS]-x(5)-[GCN]-x-[LIVMFY].

NAME: Putative AMP-binding domain signature.
CONSENSUS: [LIVMFY]-x(2)-[STG]-[STAG]-G-[ST]-[STEI]-[SG]-x-[PASLIVM]-[KR].

NAME: Molybdenum cofactor biosynthesis proteins signature 1.
CONSENSUS: [LIVM](3)-[LIT](2)-G-G-T-G-x(4)-D.

NAME: Molybdenum cofactor biosynthesis proteins signature 2.
CONSENSUS: S-x-[GS]-x(2)-D-x(5)-[LIVW]-x(10,12)-[LIV]-x(2)-[KR]-P-G-[KRL]-P-x(2)-
CONSENSUS: [LIVMF]-[GA].

NAME: moaA / nifB / pqqE family signature.
CONSENSUS: [LIV]-x(3)-C-[NP]-[LIVMF]-[QRS]-C-x-[FYM]-C.

NAME: Radical activating enzymes signature.
CONSENSUS: [GV]-x-G-x-[KR]-x(3)-F-x(2)-G-x(0,1)-C-x(3)-C-x(2)-C-x-[NL].

NAME: Tpx family signature.
CONSENSUS: S-x-D-L-P-F-A-x(2)-[KR]-[FW]-C.

NAME: Cytochrome c family heme-binding site signature.
CONSENSUS: C-{CPWHF}-{CPWR}-C-H-{CFYW}.

NAME: Cytochrome b5 family, heme-binding domain signature.
CONSENSUS: [FY]-[LIVMK]-x(2)-H-P-[GA]-G.

NAME: Cytochrome b/b6 heme-ligand signature.
CONSENSUS: [DENQ]-x(3)-G-[FYWMQ]-x-[LIVMF]-R-x(2)-H.

NAME: Cytochrome b/b6 Qo site signature.
CONSENSUS: P-[DE]-W-[FY]-[LFY](2).

NAME: Cytochrome b559 subunits heme-binding site signature.
CONSENSUS: [LIV]-x-[ST]-[LIVF]-R-[FYW]-x(2)-[IV]-H-[STGA]-[LIV]-[STGA]-[IV]-P.

NAME: Nickel-dependent hydrogenases b-type cytochrome subunit signature 1.
CONSENSUS: R-[LIVMFYW]-x-H-W-[LIVM]-x(2)-[LIVMF]-[STAC]-[LIVM]-x(2)-L-x-[LIVM]-T-G.

NAME: Nickel-dependent hydrogenases b-type cytochrome subunit signature 2.
CONSENSUS: [RH]-[STA]-[LIVMFYW]-H-[RH]-[LIVM]-x(2)-W-x-[LIVMF]-x(2)-F-x(3)-H.

NAME: Succinate dehydrogenase cytochrome b subunit signature 1.
CONSENSUS: R-P-[LIVMT]-x(3)-[LIVM]-x(6)-[LIVMWPK]-x(4)-S-x(2)-H-R-x-[ST].

NAME: Succinate dehydrogenase cytochrome b subunit signature 2.
CONSENSUS: H-x(3)-[GA]-[LIVMT]-R-[HF]-[LIVMF]-x-[FYWM]-D-x-[GVA].

NAME: Thioredoxin family active site.
CONSENSUS: [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTN]-C-[GATPLVE]-
CONSENSUS: [PHYWSTA]-C-x(6)-[LIVMFYWT].

NAME: Glutaredoxin active site.
CONSENSUS: [LIVD]-[FYSA]-x(4)-C-[PV]-[FYW]-C-x(2)-[TAV]-x(2,3)-[LIV].

NAME: Type-1 copper (blue) proteins signature.
CONSENSUS: [GA]-x(0,2)-[YSA]-x(0,1)-[VFY]-x-C-x(1,2)-[PG]-x(0,1)-H-x(2,4)-[MQ].

NAME: 2Fe-2S ferredoxins, iron-sulfur binding region signature.
CONSENSUS: C-{C}-{C}-[GA]-{C}-C-[GAST]-{CPDEKRHFYW}-C.

NAME: Adrenodoxin family, iron-sulfur binding region signature.
CONSENSUS: C-x(2)-[STAQ]-x-[STAMV]-C-[STA]-T-C-[HR].

NAME: 4Fe-4S ferredoxins, iron-sulfur binding region signature.
CONSENSUS: C-x(2)-C-x(2)-C-x(3)-C-[PEG].

NAME: High potential iron-sulfur proteins signature.
CONSENSUS: C-x(6,9)-[LIVM]-x(3)-G-[YW]-C-x(2)-[FYW].

NAME: Rieske iron-sulfur protein signature 1.
CONSENSUS: C-[TK]-H-L-G-C-[LIVT].

NAME: Rieske iron-sulfur protein signature 2.
CONSENSUS: C-P-C-H-x-[GSA].

NAME: Flavodoxin signature.
CONSENSUS: [LIV]-[LIVFY]-[FY]-x-[ST]-x(2)-[AGC]-x-T-x(3)-A-x(2)-[LIV].

NAME: Rubredoxin signature.
CONSENSUS: [LIVM]-x(3)-W-x-C-P-x-C-[AGD].

NAME: Electron transfer flavoprotein alpha-subunit signature.
CONSENSUS: [LI]-Y-[LIVM]-[AT]-x-G-[IV]-[SD]-G-x-[IV]-Q-H-x(2)-G-x(6)-[IV]-x-A-
CONSENSUS: [IV]-N.

NAME: Electron transfer flavoprotein beta-subunit signature.
CONSENSUS: [IVA]-x-[KR]-x(2)-[DE]-[GD]-[GDE]-x(1,2)-[EQ]-x-[LIV]-x(4)-P-x-[LIVM](2)-
CONSENSUS: [TAC].

NAME: Vertebrate metallothioneins signature.
CONSENSUS: C-x-C-[GSTAP]-x(2)-C-x-C-x(2)-C-x-C-x(2)-C-x-K.

NAME: Ferritin iron-binding regions signature 1.
CONSENSUS: E-x-[KR]-E-x(2)-E-[KR]-[LF]-[LIVMA]-x(2)-Q-N-x-R-x-G-R.

NAME: Ferritin iron-binding regions signature 2.
CONSENSUS: D-x(2)-[LIVMF]-[STAC]-[DH]-F-[LI]-[EN]-x(2)-[FY]-L-x(6)-[LIVM]-[KN].

NAME: Bacterioferritin signature.
CONSENSUS: <M-x-G-x(3)-V-[LIV]-x(2)-[LM]-x(3)-L-x(3)-L.

NAME: Transferrins signature 1.
CONSENSUS: Y-x(0,1)-[VAS]-V-[IVAC]-[IVA]-[IVA]-[RKH]-[RKS]-[GDENSA].

NAME: Transferrins signature 2.
CONSENSUS: Y-x-G-A-[FL]-[KRHNQ]-C-L-x(3,4)-G-[DENQ]-V-[GA]-[FYW].

NAME: Transferrins signature 3.
CONSENSUS: [DENQ]-[YF]-x-[LY]-L-C-x-[DN]-x(5,8)-[LIV]-x(4,5)-C-x(2)-A-x(4)-[HQR]-x-
CONSENSUS: [LIVMFYW]-[LIVM].

NAME: Globins profile.

NAME: Protozoan/cyanobacterial globins signature.
CONSENSUS: F-[LF]-x(5)-G-[PA]-x(4)-G-[KRA]-x-[LIVM]-x(3)-H.

NAME: Plant hemoglobins signature.
CONSENSUS: [SN]-P-x-L-x(2)-H-A-x(3)-F.

NAME: Hemerythrins signature.
CONSENSUS: W-L-x-[NQ]-H-I-x(3)-D-F.

NAME: Arthropod hemocyanins / insect LSPs signature 1.
CONSENSUS: Y-[FYW]-x-E-D-[LIVM]-x(2)-N-x(6)-H-x(3)-P.

NAME: Arthropod hemocyanins / insect LSPs signature 2.
CONSENSUS: T-x(2)-R-D-P-x-[FY]-[FYW].

NAME: Heavy-metal-associated domain.
CONSENSUS: [LIVN]-x(2)-[LIVMFA]-x-C-x-[STAGCDNH]-C-x(3)-[LIVFG]-x(3)-[LIV]-x(9,11)-
CONSENSUS: [IVA]-x-[LVFYS].

NAME: ABC transporters family signature.
CONSENSUS: [LIVMFYC]-[SA]-[SAPGLVFYKQH]-G-[DENQMW]-[KRQASPCLIMFW]-[KRNQSTAVM]-
CONSENSUS: [KRACLVM]-[LIVMFYPAN]-{PHY}-[LIVMFW]-[SAGCLIVP]-{FYWHP}-{KRHP}-
CONSENSUS: [LIVMFYWSTA].

NAME: Binding-protein-dependent transport systems inner membrane comp. sign.
CONSENSUS: [LIVMFY]-x(8)-[EQR]-[STAGV]-[STAG]-x(3)-G-[LIVMFYSTAC]-x(5)-[LIVMFYSTA]-
CONSENSUS: x(4)-[LIVMFY]-[PKR].

NAME: ABC-2 type transport system integral membrane proteins signature.
CONSENSUS: [LIMST]-x(2)-[LIMW]-x(2)-[LIMCA]-[GSTC]-x-[GSAIV]-x(6)-[LIMGA]-[PGSNQ]-
CONSENSUS: x(9,12)-P-[LIMFT]-x-[HRSY]-x(5)-[RQ].

NAME: Bacterial extracellular solute-binding proteins, family 1 signature.
CONSENSUS: [GAP]-[LIVMFA]-[STAVDN]-x(4)-[GSAV]-[LIVMFY](2)-Y-[ND]-x(3)-[LIVMF]-x-
CONSENSUS: [KNDE].

NAME: Bacterial extracellular solute-binding proteins, family 3 signature.
CONSENSUS: G-[FYIL]-[DE]-[LIVMT]-[DE]-[LIVMF]-x(3)-[LIVMA]-[VAGC]-x(2)-[LIVMAGN].

NAME: Bacterial extracellular solute-binding proteins, family 5 signature.
CONSENSUS: [AG]-x(6,7)-[DNEG]-x(2)-[STAVE]-[LIVMFYWA]-x-[LIVMFY]-x-[LIVM]-[KR]-

CONSENSUS: [KRHDE]-[GDN]-[LIVMA]-[KNGSP]-[FW].

NAME: Serum albumin family signature.
CONSENSUS: [FY]-x(6)-C-C-x(7)-C-[LFY]-x(6)-[LIVMFYW].

NAME: Transthyretin signature 1.
CONSENSUS: S-K-C-P-L-M-V-K-V-L-D-[AS]-V-R-G.

NAME: Transthyretin signature 2.
CONSENSUS: S-P-[FY]-S-[FY]-S-T-T-A-[LIVM]-V-[ST]-x-P.

NAME: Avidin / Streptavidin family signature.
CONSENSUS: [DEN]-x(2)-[KR]-[STA]-x(2)-V-G-x-[DN]-x-[FW]-T-[KR].

NAME: Eukaryotic cobalamin-binding proteins signature.
CONSENSUS: [SN]-V-D-T-[GA]-A-[LIVM]-A-x-L-A-[LIVMF]-T-C.

NAME: Lipocalin signature.
CONSENSUS: [DENG]-x-[DENQGSTARK]-x(0,2)-[DENQARK]-[LIVFY]-{CP}-G-{C}-W-[FYWLRH]-x-
CONSENSUS: [LIVMTA].

NAME: Cytosolic fatty-acid binding proteins signature.
CONSENSUS: [GSAIVK]-x-[FYW]-x-[LIVMF]-x(4)-[NHG]-[FY]-[DE]-x-[LIVMFY]-[LIVM]-x(2)-
CONSENSUS: [LIVMAKR].

NAME: Acyl-CoA-binding protein signature.
CONSENSUS: P-[STA]-x-[DEN]-x-[LIVMF]-x(2)-[LIVMFY]-Y-[GSTA]-x-[FY]-K-Q-[STA](2)-x-G.

NAME: LBP / BPI / CETP family signature.
CONSENSUS: [PA]-[GA]-[LIVMC]-x(2)-R-[IV]-[ST]-x(3)-L-x(5)-[EQ]-x(4)-[LIVM]-[EQK]-
CONSENSUS: x(8)-P.

NAME: Phosphatidylethanolamine-binding protein family signature.
CONSENSUS: [FY]-x-[LIVMF](3)-x-[DC]-P-D-x-P-[SN]-x(10)-H.

NAME: Plant lipid transfer proteins signature.
CONSENSUS: [LIVM]-[PA]-x(2)-C-x-[LIVM]-x-[LIVM]-x-[LIVMFY]-x-[LIVM]-[ST]-x(3)-
CONSENSUS: [DN]-C-x(2)-[LIVM].

NAME: Uteroglobin family signature 1.
CONSENSUS: [GA]-x(3)-I-C-P-x-[LIVMF]-x(3)-[LIVM]-[DE]-x-[LIVMF](2).

NAME: Uteroglobin family signature 2.
CONSENSUS: [DEQ]-x(4)-[SN]-x(5)-[DEQ]-x-I-x(2)-S-[PSE]-[LS]-C.

NAME: Mitochondrial energy transfer proteins signature.
CONSENSUS: P-x-[DE]-x-[LIVAT]-[RK]-x-[LRH]-[LIVMFY]-[QMAIGV].

NAME: Sugar transport proteins signature 1.
CONSENSUS: [LIVMSTAG]-[LIVMFSAG]-x(2)-[LIVMSA]-[DE]-x-[LIVMFYWA]-G-R-[RK]-x(4,6)-
CONSENSUS: [GSTA].

NAME: Sugar transport proteins signature 2.
CONSENSUS: [LIVMF]-x-G-[LIVMFA]-x(2)-G-x(8)-[LIFY]-x(2)-[EQ]-x(6)-[RK].

NAME: LacY family proton/sugar symporters signature 1.
CONSENSUS: G-[LIVM](2)-x-D-[RK]-L-G-L-[RK](2)-x-[LIVM](2)-W.

NAME: LacY family proton/sugar symporters signature 2.
CONSENSUS: P-x-[LIVMF](2)-N-R-[LIVM]-G-x-K-N-[STA]-[LIVM](3).

NAME: PTR2 family proton/oligopeptide symporters signature 1.
CONSENSUS: [GA]-[GAS]-[LIVMFYWA]-[LIVM]-[GAS]-D-x-[LIVMFYWT]-[LIVMFYW]-G-x(3)-[TAV]-
CONSENSUS: [IV]-x(3)-[GSTAV]-x-[LIVMF]-x(3)-[GA].

NAME: PTR2 family proton/oligopeptide symporters signature 2.
CONSENSUS: [FYT]-x(2)-[LMFY]-[FYV]-[LIVMFYWA]-x-[IVG]-N-[LIVMAG]-G-[GSA]-[LIMF].

NAME: Amiloride-sensitive sodium channels signature.
CONSENSUS: Y-x(2)-[EQTF]-x-C-x(2)-[GSTDNL]-C-x-[QT]-x(2)-[LIVMT]-[LIVMS]-x(2)-C-x-C.

NAME: Sodium:alanine symporter family signature.
CONSENSUS: G-G-x-[GA](2)-[LIVM]-F-W-M-W-[LIVM]-x-[STAV]-[LIVMFA](2)-G.

NAME: Sodium:dicarboxylate symporter family signature 1.
CONSENSUS: P-x(0,1)-G-[DE]-x-[LIVMF](2)-x-[LIVM](2)-[KREQ]-[LIVM](3)-x-P.

NAME: Sodium:dicarboxylate symporter family signature 2.
CONSENSUS: P-x-G-x-[STA]-x-[NT]-[LIVMC]-D-G-[STAN]-x-[LIVM]-[FY]-x(2)-[LIVM]-x(2)-
CONSENSUS: [LIVM]-[FY]-[LI]-[SA]-Q.

NAME: Sodium:galactoside symporter family signature.
CONSENSUS: D-x(3)-G-x(3)-[DN]-x(6,8)-G-[KH]-F-[KR]-P-[FYW]-[LIVM](2)-x-[GSTA](2).

NAME: Sodium:neurotransmitter symporter family signature 1.
CONSENSUS: W-R-F-[GP]-Y-x(4)-N-G-G-G-x-[FY].

NAME: Sodium:neurotransmitter symporter family signature 2.
CONSENSUS: Y-[LIVMFY]-x(2)-[SC]-[LIVMFY]-[STQ]-x(2)-L-P-W-x(2)-C-x(4)-N-[GST].

NAME: Sodium:solute symporter family signature 1.
CONSENSUS: [GS]-x(2)-[LIY]-x(3)-[LIVMFYWSTAG](10)-[LIY]-[TAV]-x(2)-G-G-[LMF]-x-
CONSENSUS: [SAP].

NAME: Sodium:solute symporter family signature 2.
CONSENSUS: [GAST]-[LIVM]-x(3)-[KR]-x(4)-G-A-x(2)-[GAS]-[LIVMGS]-[LIVMW]-[LIVMGAT]-G-
CONSENSUS: x-[LIVMG].

NAME: Sodium:sulfate symporter family signature.
CONSENSUS: [STACP]-S-x(2)-F-x(2)-P-[LIVM]-[GSA]-x(3)-N-x-[LIVM]-V.

NAME: glpT family of transporters signature.
CONSENSUS: R-G-x(5)-W-N-x(2)-H-N-x-G-G.

NAME: Ammonium transporters signature.
CONSENSUS: D-[FYWS]-A-G-[GSC]-x(2)-[IV]-x(3)-[SAG](2)-x(2)-[SAG]-[LIVMF]-x(3)-
CONSENSUS: [LIVMFYWA](2)-x-[GK]-x-R.

NAME: BCCT family of transporters signature.
CONSENSUS: [GSDN]-W-T-[LIVM]-x-[FY]-W-x-W-W.

NAME: Flagellar motor protein motA family signature.
CONSENSUS: A-[LMF]-x-[GAT]-T-[LIVF]-x-G-x-[LIVMF]-x(7)-P.

NAME: Formate and nitrite transporters signature 1.
CONSENSUS: [LIVMA]-[LIVMY]-x-G-[GSTA]-[DES]-L-[FI]-[TN]-[GS].

NAME: Formate and nitrite transporters signature 2.
CONSENSUS: [GA]-x(2)-[CA]-N-[LIVMFYW](2)-V-C-[LV]-A.

NAME: Prokaryotic sulfate-binding proteins signature 1.
CONSENSUS: K-x-[NQEK]-[GT]-G-[DQ]-x-[LIVM]-x(3)-Q-S.

NAME: Prokaryotic sulfate-binding proteins signature 2.
CONSENSUS: N-P-K-[ST]-S-G-x-A-R.

NAME: Sulfate transporters signature.
CONSENSUS: P-x-Y-[GS]-L-Y-[STAG](2)-x(4)-[LIVMFY](3)-x(3)-[GSTA](2)-S-[KR].

NAME: Amino acid permeases signature.
CONSENSUS: [STAGC]-G-[PAG]-x(2,3)-[LIVMFYWA](2)-x-[LIVMFYW]-x-[LIVMFWSTAGC](2)-
CONSENSUS: [STAGC]-x(3)-[LIVMFYW]-x-[LIVMST]-x(3)-[LIMCTA]-[GA]-E-x(5)-[PSAL].

NAME: Aromatic amino acids permeases signature.
CONSENSUS: I-G-[GA]-G-M-[LF]-[SA]-x-P-x(3)-[SA]-G-x(2)-F.

NAME: Xanthine/uracil permeases family signature.
CONSENSUS: [LIVM]-P-x-[PASIF]-V-[LIVM]-G-G-x(4)-[LIVM]-[FY]-[GSA]-x-[LIVM]-x(3)-G.

NAME: Anion exchangers family signature 1.
CONSENSUS: F-G-G-[LIVM](2)-[KR]-D-[LIVM]-[RK]-R-R-Y.

NAME: Anion exchangers family signature 2.
CONSENSUS: [FI]-L-I-S-L-I-F-I-Y-E-T-F-x-K-L.

NAME: MIP family signature.
CONSENSUS: [HNQA]-x-N-P-[STA]-[LIVMF]-[ST]-[LIVMF]-[GSTAFY].

NAME: General diffusion Gram-negative porins signature.
CONSENSUS: [LIVMFY]-x(2)-G-x(2)-Y-x-F-x-K-x(2)-[SN]-[STAV]-[LIVMFYW]-V.

NAME: OmpA-like domain.
CONSENSUS: [LIVMA]-x-[GT]-x-[TA]-[DA]-x(2)-[DG]-[GSTP]-x(2)-[LFYDE]-[NQS]-x(2)-
CONSENSUS: [LI]-[SG]-[QE]-[KRQE]-R-A-x(2)-[LV]-x(3)-[LIVMF]-x(4,5)-[LIVM]-x(4)-
CONSENSUS: [LIVM]-x(3)-[SG]-x-G.

NAME: Eukaryotic mitochondrial porin signature.
CONSENSUS: [YH]-x(2)-D-[SPA]-x-[STA]-x(3)-[TAG]-[KR]-[LIVMF]-[DNSTA]-[DNS]-x(4)-
CONSENSUS: [GSTAN]-[LIVMA]-x-[LIVMY].

NAME: Insulin-like growth factor binding proteins signature.
CONSENSUS: G-C-[GS]-C-C-x(2)-C-A-x(6)-C.

NAME: GPR1/FUN34/yaaH family signature.
CONSENSUS: N-P-[AV]-P-[LF]-G-L-x-[GSA]-F.

NAME: GNS1/SUR4 family signature.
CONSENSUS: L-x-F-L-H-x-Y-H-H.

NAME: 43 Kd postsynaptic protein signature.
CONSENSUS: G-Q-D-Q-T-K-Q-Q-I.

NAME: Actins signature 1.
CONSENSUS: [FY]-[LIV]-G-[DE]-E-A-Q-x-[RKQ](2)-G.

NAME: Actins signature 2.
CONSENSUS: W-[IV]-[STA]-[RK]-x-[DE]-Y-[DNE]-[DE].

NAME: Actins and actin-related proteins signature.
CONSENSUS: [LM]-[LIVM]-T-E-[GAPQ]-x-[LIVMFYWHQ]-N-[PSTAQ]-x(2)-N-[KR].

NAME: Annexins repeated domain signature.
CONSENSUS: [TG]-[STV]-x(8)-[LIVMF]-x(2)-R-x(3)-[DEQNH]-x(7)-[IFY]-x(7)-[LIVMF]-
CONSENSUS: x(3)-[LIVMF]-x(11)-[LIVMFA]-x(2)-[LIVMF].

NAME: Caveolins signature.
CONSENSUS: F-E-D-V-I-A-E-P.

NAME: Clathrin light chain signature 1.
CONSENSUS: F-L-A-Q-Q-E-S.

NAME: Clathrin light chain signature 2.
CONSENSUS: [KR]-D-x-S-[KR]-[LIVM]-[KR]-x-[LIVM](3)-x-L-K.

NAME: Clusterin signature 1.
CONSENSUS: C-K-P-C-L-K-x-T-C.

NAME: Clusterin signature 2.
CONSENSUS: C-L-[RK]-M-[RK]-x-[EQ]-C-[ED]-K-C.

NAME: Connexins signature 1.
CONSENSUS: C-[DN]-T-x-Q-P-G-C-x(2)-V-C-Y-D.

NAME: Connexins signature 2.
CONSENSUS: C-x(3,4)-P-C-x(3)-[LIVM]-[DEN]-C-[FY]-[LIVM]-[SA]-[KR]-P.

NAME: Crystallins beta and gamma 'Greek key' motif signature.
CONSENSUS: [LIVMFYWA]-x-{DEHRKSTP}-[FY]-[DEQHKY]-x(3)-[FY]-x-G-x(4)-[LIVMFCST].

NAME: Dynamin family signature.
CONSENSUS: L-P-[RK]-G-[STN]-[GN]-[LIVM]-V-T-R.

NAME: Dynein light chain type 1 signature.
CONSENSUS: H-x-I-x-G-[KR]-x-F-[GA]-S-x-V-[ST]-[HY]-E.

NAME: FtsZ protein signature 1.
CONSENSUS: N-[ST]-D-x-Q-x-L-x(16,18)-G-x-G-[ATV]-G-[GSAN]-x-P-x(2)-G.

NAME: FtsZ protein signature 2.
CONSENSUS: [DNHKR]-[LIVMF]-x-[LIVMF](2)-[VSTAC]-[STAC]-G-x-G-[GK]-G-T-G-[ST]-G-
CONSENSUS: [GSAR]-[STA]-P-[LIVMFT]-[LIVMF]-[SGAV].

NAME: Fungal hydrophobins signature.
CONSENSUS: [GN]-[DNQPSA]-x-C-[GSTANK]-[GSTADNQ]-[STNQI]-[PTIV]-x-C-C-[DENQKPST].

NAME: Intermediate filaments signature.
CONSENSUS: [IV]-x-[TACI]-Y-[RKH]-x-[LM]-L-[DE].

NAME: Involucrin signature.
CONSENSUS: <M-S-[QH]-Q-x-T-[LV]-P-V-T-[LV].

NAME: Kinesin motor domain signature.
CONSENSUS: [GSA]-[KRHPSTQVM]-[LIVMF]-x-[LIVMF]-[IVC]-D-L-[AH]-G-[SAN]-E.

NAME: Kinesin motor domain profile.

NAME: Kinesin light chain repeat.
CONSENSUS: [DEQR]-A-L-x(3)-[GEQ]-x(3)-G-x-[DNS]-x-P-x-V-A-x(3)-N-x-L-[AS]-
CONSENSUS: x(5)-[QR]-x-[KR]-[FY]-x(2)-[AV]-x(4)-[HKNQ].

NAME: Myelin basic protein signature.
CONSENSUS: V-V-H-F-F-K-N.

NAME: Myelin P0 protein signature.
CONSENSUS: S-[KR]-S-x-K-[AG]-x-[SA]-E-K-K-[STA]-K.

NAME: Myelin proteolipid protein signature 1.
CONSENSUS: G-[MV]-A-L-F-C-G-C-G-H.

NAME: Myelin proteolipid protein signature 2.
CONSENSUS: C-x-[ST]-x-[DE]-x(3)-[ST]-[FY]-x-L-[FY]-I-x(4)-G-A.

NAME: Neuromodulin (GAP-43) signature 1.
CONSENSUS: <M-L-C-C-[LIVM]-R-R.

NAME: Neuromodulin (GAP-43) signature 2.
CONSENSUS: S-F-R-G-H-I-x-R-K-K-[LIVM].

NAME: Osteopontin signature.
CONSENSUS: [KQ]-x-[TA]-x(2)-[GA]-S-S-E-E-K.

NAME: Peripherin / rom-1 signature.
CONSENSUS: D-[GS]-V-P-F-[ST]-C-C-N-P-x-S-P-R-P-C.

NAME: Profilin signature.
CONSENSUS: <x(0,1)-[STA]-x(0,1)-W-[DENQH]-x-[YI]-x-[DEQ].

NAME: Surfactant associated polypeptide SP-C palmitoylation sites.
CONSENSUS: I-P-C-C-P-V.

NAME: Synapsins signature 1.
CONSENSUS: L-R-R-R-L-S-D-S.

NAME: Synapsins signature 2.
CONSENSUS: G-H-A-H-S-G-M-G-K-V-K.

NAME: Synaptobrevin signature.
CONSENSUS: N-[LIVM]-[DENS]-[KL]-V-x-[DEQ]-R-x(2)-[KR]-[LIVM]-[STDE]-x-[LIVM]-x-[DE]-
CONSENSUS: [KR]-[TA]-[DE].

NAME: Synaptophysin / synaptoporin signature.
CONSENSUS: L-S-V-[DE]-C-x-N-K-T.

NAME: Tropomyosins signature.
CONSENSUS: L-K-E-A-E-x-R-A-E.

NAME: Tubulin subunits alpha, beta, and gamma signature.
CONSENSUS: [SAG]-G-G-T-G-[SA]-G.

NAME: Tubulin-beta mRNA autoregulation signal.
CONSENSUS: <M-R-[DE]-[IL].

NAME: Tau and MAP proteins tubulin-binding domain signature.
CONSENSUS: G-S-x(2)-N-x(2)-H-x-[PA]-[AG]-G(2).

NAME: Neuraxin and MAP1B proteins repeated region signature.

CONSENSUS: [STAGDN]-Y-x-Y-E-x(2)-[DE]-[KR]-[STAGCI].

NAME: F-actin capping protein alpha subunit signature 1.
CONSENSUS: V-H-[FY](2)-E-D-G-N-V.

NAME: F-actin capping protein alpha subunit signature 2.
CONSENSUS: F-K-[AE]-L-R-R-x-L-P.

NAME: F-actin capping protein beta subunit signature.
CONSENSUS: C-D-Y-N-R-D.

NAME: Vinculin family talin-binding region signature.
CONSENSUS: [KR]-x-[LIVMF]-x(3)-[LIVMA]-x(2)-[LIVM]-x(6)-R-Q-Q-E-L.

NAME: Vinculin repeated domain signature.
CONSENSUS: [LIVM]-x-[QA]-A-x(2)-W-[IL]-x-[DN]-P.

NAME: Amyloidogenic glycoprotein extracellular domain signature.
CONSENSUS: G-[VT]-E-[FY]-V-C-C-P.

NAME: Amyloidogenic glycoprotein intracellular domain signature.
CONSENSUS: G-Y-E-N-P-T-Y-[KR].

NAME: Cadherins extracellular repeated domain signature.
CONSENSUS: [LIV]-x-[LIV]-x-D-x-N-D-[NH]-x-P.

NAME: Insect cuticle proteins signature.
CONSENSUS: G-x(7)-[DEN]-G-x(6)-Y-x-A-[DNG]-x(2,3)-G-[FY]-x-[AP].

NAME: Gas vesicles protein GVPa signature 1.
CONSENSUS: [LIVM]-x-[DE]-[LIVMFYT]-[LIVM]-[DE]-x-[LIVM](2)-[DKR](2)-G-x-[LIVM](2).

NAME: Gas vesicles protein GVPa signature 2.
CONSENSUS: R-[LIVA](3)-A-[GS]-[LIVMFY]-x-T-x(3)-Y-[AG].

NAME: Gas vesicles protein GVPc repeated domain signature.
CONSENSUS: F-L-x(2)-T-x(3)-R-x(3)-A-x(2)-Q-x(3)-L-x(2)-F.

NAME: Bacterial microcompartiments proteins signature.
CONSENSUS: D-x(0,1)-M-x-K-[SAG](2)-x-[IV]-x-[LIVM]-[LIVMA]-[GCS]-x(4)-[GD]-[SGPD]-
CONSENSUS: [GA].

NAME: Flagella basal body rod proteins signature.
CONSENSUS: [GTARYQ]-x(9)-[LIVMYSTA](2)-[GSTA]-[STADEN]-N-[LIVM]-[SAN]-N-x-[SADNFR]-
CONSENSUS: [STV].

NAME: Flagella transport protein fliP family signature 1.
CONSENSUS: [PA]-A-[FY]-x-[LIVT]-[STH]-[EQ]-[LI]-x(2)-[GA]-F-[KREQ]-[IM]-G-[LIF].

NAME: Flagella transport protein fliP family signature 2.
CONSENSUS: P-[LIVMF]-K-[LIVMF](5)-x-[LIVMA]-[DNGS]-G-W.

NAME: Plant viruses icosahedral capsid proteins 'S' region signature.
CONSENSUS: [FYW]-x-[PSTA]-x(7)-G-x-[LIVM]-x-[LIVM]-x-[FYWI]-x(2)-D-x(5)-P.

NAME: Potexviruses and carlaviruses coat protein signature.
CONSENSUS: [RK]-[FYW]-A-[GAP]-F-D-x-F-x(2)-[LV]-x(3)-[GAST](2).

NAME: Neurotransmitter-gated ion-channels signature.
CONSENSUS: C-x-[LIVMFQ]-x-[LIVMF]-x(2)-[FY]-P-x-D-x(3)-C.

NAME: ATP P2X receptors signature.
CONSENSUS: G-G-x-[LIVM]-G-[LIVM]-x-[IV]-x-W-x-C-[DN]-L-D-x(5)-C-x-P-x-Y-x-F.

NAME: G-protein coupled receptors signature.
CONSENSUS: [GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)-[LIVMFT]-
CONSENSUS: [GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)-[LIVM].

NAME: G-protein coupled receptors family 2 signature 1.
CONSENSUS: C-x(3)-[FYWLIV]-D-x(3,4)-C-[FW]-x(2)-[STAGV]-x(8,9)-C-[PF].

NAME: G-protein coupled receptors family 2 signature 2.
CONSENSUS: Q-G-[LMFCA]-[LIVMFT]-[LIV]-x-[LIVFST]-[LIF]-[VFYH]-C-[LFY]-x-N-x(2)-V.

NAME: G-protein coupled receptors family 3 signature 1.
CONSENSUS: [LV]-x-N-[LIVM](2)-x-L-F-x-I-[PA]-Q-[LIVM]-[STA]-x-[STA](3)-[STAN].

NAME: G-protein coupled receptors family 3 signature 2.
CONSENSUS: C-C-[FYW]-x-C-x(2)-C-x(4)-[FYW]-x(2,4)-[DN]-x(2)-[STAH]-C-x(2)-C.

NAME: G-protein coupled receptors family 3 signature 3.
CONSENSUS: F-N-E-[STA]-K-x-I-[STAG]-F-[ST]-M.

NAME: Visual pigments (opsins) retinal binding site.
CONSENSUS: [LIVMWAC]-[PGAC]-x(3)-[SAC]-K-[STALIMR]-[GSACPNV]-[STACP]-x(2)-[DENF]-
CONSENSUS: [AP]-x(2)-[IY].

NAME: Bacterial rhodopsins signature 1.
CONSENSUS: R-Y-x-[DT]-W-x-[LIVMF]-[ST]-T-P-[LIVM](3).

NAME: Bacterial rhodopsins retinal binding site.
CONSENSUS: [FYIV]-x-[FYVG]-[LIVM]-D-[LIVMF]-x-[STA]-K-x(2)-[FY].

NAME: Receptor tyrosine kinase class II signature.
CONSENSUS: [DN]-[LIV]-Y-x(3)-Y-Y-R.

NAME: Receptor tyrosine kinase class III signature.
CONSENSUS: G-x-H-x-N-[LIVM]-V-N-L-L-G-A-C-T.

NAME: Receptor tyrosine kinase class V signature 1.
CONSENSUS: F-x-[DN]-x-[GAW]-[GA]-C-[LIVM]-[SA]-[LIVM](2)-[SA]-[LV]-[KRHQ]-[LIVA]-
CONSENSUS: x(3)-[KR]-C-[PSAW].

NAME: Receptor tyrosine kinase class V signature 2.
CONSENSUS: C-x(2)-[DE]-G-[DEQ]-W-x(2,3)-[PAQ]-[LIVMT]-[GT]-x-C-x-C-x(2)-G-[HFY]-
CONSENSUS: [EQ].

NAME: Growth factor and cytokines receptors family signature 1.
CONSENSUS: C-[LVFYR]-x(7,8)-[STIVDN]-C-x-W.

NAME: Growth factor and cytokines receptors family signature 2.
CONSENSUS: [STGL]-x-W-[SG]-x-W-S.

NAME: TNFR/NGFR family cysteine-rich region signature.
CONSENSUS: C-x(4,6)-[FYH]-x(5,10)-C-x(0,2)-C-x(2,3)-C-x(7,11)-C-x(4,6)-[DNEQSKP]-
CONSENSUS: x(2)-C.

NAME: TNFR/NGFR family cysteine-rich region domain.

NAME: Integrins alpha chain signature.
CONSENSUS: [FYWS]-[RK]-x-G-F-F-x-R.

NAME: Integrins beta chain cysteine-rich domain signature.
CONSENSUS: C-x-[GNQ]-x(1,3)-G-x-C-x-C-x(2)-C-x-C.

NAME: Natriuretic peptides receptors signature.
CONSENSUS: G-P-x-C-x-Y-x-A-A-x-V-x-R-x(3)-H-W.

NAME: Photosynthetic reaction center proteins signature.
CONSENSUS: [NH]-x(4)-P-x-H-x(2)-[SAG]-x(11)-[SAGC]-x-H-[SAG](2).

NAME: Antenna complexes alpha subunits signature.
CONSENSUS: [LIVFAG]-x-[GASV]-[LIVFA]-x-[IV]-H-x(3)-[LIVM]-[GSTAE]-[STANH]-x(1,3)-
CONSENSUS: [STN]-W-[LIVMFYW].

NAME: Antenna complexes beta subunits signature.
CONSENSUS: [EQ]-x(4)-H-x(5)-[GSTA]-x(3)-[FY]-x(3)-[AG]-x(2)-[AV]-H-x(7)-P.

NAME: Photosystem I psaA and psaB proteins signature.
CONSENSUS: C-D-G-P-G-R-G-G-T-C.

NAME: Photosystem I psaG and psaK proteins signature.
CONSENSUS: G-F-x-[LIVM]-x-[DEA]-x(2)-[GA]-x-[GTA]-[SA]-x-G-H-x-[LIVM]-[GA].

NAME: Phytochrome chromophore attachment site signature.
CONSENSUS: [RGS]-[GSA]-[PV]-H-x-C-H-x(2)-Y.

NAME: Phytochrome chromophore attachment site domain profile.

NAME: Speract receptor repeated domain signature.
CONSENSUS: G-x(5)-G-x(2)-E-x(6)-W-G-x(2)-C-x(3)-[FYW]-x(8)-C-x(3)-G.

NAME: TonB-dependent receptor proteins signature 1.
CONSENSUS: <x(10,115)-[DENF]-[ST]-[LIVMF]-[LIVSTEQ]-V-x-[AGP]-[STANEQPK].

NAME: TonB-dependent receptor proteins signature 2.
CONSENSUS: [LYGSTANE]-x(3)-[GSTAENQ]-x-[PGE]-R-x-[LIVFYWA]-x-[LIVMFTA]-[STAGNQ]-
CONSENSUS: [LIVMFYGTA]-x-[LIVMFYWGTADQ]-x-F>.

NAME: Transmembrane 4 family signature.
CONSENSUS: G-x(3)-[LIVMF]-x(2)-[GSA]-[LIVMF](2)-G-C-x-[GA]-[STA]-x(2)-[EG]-x(2)-
CONSENSUS: [CWN]-[LIVM](2).

NAME: Bacterial chemotaxis sensory transducers signature.
CONSENSUS: R-T-E-[EQ]-Q-x(2)-[SA]-[LIVM]-x-[EQ]-T-A-A-S-M-E-Q-L-T-A-T-V.

NAME: ER lumen protein retaining receptor signature 1.
CONSENSUS: G-I-S-x-[KR]-x-Q-x-L-[FY]-x-[LIV](2)-F-x(2)-R-Y.

NAME: ER lumen protein retaining receptor signature 2.
CONSENSUS: L-E-[SA]-V-A-I-[LM]-P-Q-L.

NAME: Ephrins signature.
CONSENSUS: [KRQ]-[LF]-[CST]-x-K-[IF]-Q-x-[FY]-[ST]-[PA]-x(3)-G-x-E-F-x(5)-[FY](2)-
CONSENSUS: x(2)-[SA].

NAME: Granulins signature.
CONSENSUS: C-x-D-x(2)-H-C-C-P-x(4)-C.

NAME: HBGF/FGF family signature.
CONSENSUS: G-x-L-x-[STAGP]-x(6,7)-[DE]-C-x-[FM]-x-E-x(6)-Y.

NAME: PTN/MK heparin-binding protein family signature 1.
CONSENSUS: S-[DE]-C-x-[DE]-W-x-W-x(2)-C-x-P-x-[SN]-x-D-C-G-[LIVMA]-G-x-R-E-G.

NAME: PTN/MK heparin-binding protein family signature 2.
CONSENSUS: C-[KR]-[LIVM]-P-C-N-W-K-K-x-F-G-A-[DE]-C-K-Y-x-F-[EQ]-x-W-G-x-C.

NAME: Nerve growth factor family signature.
CONSENSUS: G-C-[KR]-G-[LIV]-[DE]-x(3)-[YW]-x-S-x-C.

NAME: Platelet-derived growth factor (PDGF) family signature.
CONSENSUS: P-[PS]-C-V-x(3)-R-C-[GSTA]-G-C-C.

NAME: Small cytokines (intercrine/chemokine) C-x-C subfamily signature.
CONSENSUS: C-x-C-[LIVM]-x(5,6)-[LIVMFY]-x(2)-[RKSEQ]-x-[LIVM]-x(2)-[LIVM]-x(5)-
CONSENSUS: [SAG]-x(2)-C-x(3)-[EQ]-[LIVM](2)-x(9,10)-C-L-[DN].

NAME: Small cytokines (intercrine/chemokine) C-C subfamily signature.
CONSENSUS: C-C-[LIFYT]-x(5,6)-[LI]-x(4)-[LIVMF]-x(2)-[FYW]-x(6,8)-C-x(3,4)-[SAG]-
CONSENSUS: [LIVM](2)-[FL]-x(8)-C-[STA].

NAME: TGF-beta family signature.
CONSENSUS: [LIVM]-x(2)-P-x(2)-[FY]-x(4)-C-x-G-x-C.

NAME: TNF family signature.
CONSENSUS: [LV]-x-[LIVM]-x(3)-G-[LIVMF]-Y-[LIVMFY](2)-x(2)-[QEKHL]-[LIVMGT]-x-
CONSENSUS: [LIVMFY].

NAME: TNF family profile.

NAME: Wnt-1 family signature.
CONSENSUS: C-K-C-H-G-[LIVMT]-S-G-x-C.

NAME: Interferon alpha, beta and delta family signature.
CONSENSUS: [FYH]-[FY]-x-[GNRC]-[LIVM]-x(2)-[FY]-L-x(7)-[CY]-A-W.

NAME: Granulocyte-macrophage colony-stimulating factor signature.
CONSENSUS: C-P-[LP]-T-x-E-[ST]-x-C.

NAME: Interleukin-1 signature.
CONSENSUS: [FC]-x-S-[ASLV]-x(2)-P-x(2)-[FYLIV]-[LI]-[SCA]-T-x(7)-[LIVM].

NAME: Interleukin-2 signature.
CONSENSUS: T-E-[LF]-x(2)-L-x-C-L-x(2)-E-L.

NAME: Interleukins -4 and -13 signature.
CONSENSUS: L-x-E-[LIVM](2)-x(4,5)-[LIVM]-[TL]-x(5,7)-C-x(4)-[IVA]-x-[DNS]-[LIVMA].

NAME: Interleukin-6 / G-CSF / MGF signature.
CONSENSUS: C-x(9)-C-x(6)-G-L-x(2)-[FY]-x(3)-L.

NAME: Interleukin-7 and -9 signature.
CONSENSUS: N-x-[LAP]-[SCT]-F-L-K-x-L-L.

NAME: Interleukin-10 family signature.
CONSENSUS: [GS]-C-x(2)-[LV]-x(2)-[LIVM](2)-x-F-Y-L-x(2)-V.

NAME: LIF / OSM family signature.
CONSENSUS: [PST]-x(4)-F-[NQ]-x-K-x(3)-C-x-[LF]-L-x(2)-Y-[HK].

NAME: Macrophage migration inhibitory factor family signature.
CONSENSUS: [DE]-P-C-A-x(3)-[LIVM]-x-S-I-G-x-[LIVM]-G.

NAME: Adipokinetic hormone family signature.
CONSENSUS: Q-[LV]-[NT]-[FY]-[ST]-x(2)-W.

NAME: Bombesin-like peptides family signature.
CONSENSUS: W-A-x-G-[SH]-[LF]-M.

NAME: Calcitonin / CGRP / IAPP family signature.
CONSENSUS: C-[SAGDN]-[STN]-x(0,1)-[SA]-T-C-[VMA]-x(3)-[LYF]-x(3)-[LYF].

NAME: Corticotropin-releasing factor family signature.
CONSENSUS: [PQ]-x-[LIVM]-S-[LIVM]-x(2)-[PST]-[LIVMF]-x-[LIVM]-L-R-x(2)-[LIVM].

NAME: Crustacean CHH/MIH/GIH neurohormones family signature.
CONSENSUS: C-[DENK]-D-C-x-N-[LIV]-[FY]-R-x(7)-C-[KR]-x(2)-C.

NAME: Erythropoietin / thrombopoeitin signature.
CONSENSUS: P-x(4)-C-D-x-R-[LIVM](2)-x-[KR]-x(14)-C.

NAME: Granins signature 1.
CONSENSUS: [DE]-[SN]-L-[SAN]-x(2)-[DE]-x-E-L.

NAME: Granins signature 2.
CONSENSUS: C-[LIVM](2)-E-[LIVM](2)-S-[DN]-[STA]-L-x-K-x-S-x(3)-[LIVM]-[STA]-x-E-C.

NAME: Galanin signature.
CONSENSUS: G-W-T-L-N-S-A-G-Y-L-L-G-P-H.

NAME: Gastrin / cholecystokinin family signature.
CONSENSUS: Y-x(0,1)-[GD]-[WH]-M-[DR]-F.

NAME: Glucagon / GIP / secretin / VIP family signature.
CONSENSUS: [YH]-[STAIVGD]-[DEQ]-[AGF]-[LIVMSTE]-[FYLR]-x-[DENSTAK]-[DENSTA]-
CONSENSUS: [LIVMFYG]-x(9)-[KREQL]-[KRDENQL]-[LVFYWG]-[LIVQ].

NAME: Glycoprotein hormones alpha chain signature 1.
CONSENSUS: C-x-G-C-C-[FY]-S-R-A-[FY]-P-T-P.

NAME: Glycoprotein hormones alpha chain signature 2.
CONSENSUS: N-H-T-x-C-x-C-x-T-C-x(2)-H-K.

NAME: Glycoprotein hormones beta chain signature 1.
CONSENSUS: C-[STAGM]-G-[HFYL]-C-x-[ST].

NAME: Glycoprotein hormones beta chain signature 2.
CONSENSUS: [PA]-V-A-x(2)-C-x-C-x(2)-C-x(4)-[STD]-[DEY]-C-x(6,8)-[PGSTAVM]-x(2)-C.

NAME: Gonadotropin-releasing hormones signature.
CONSENSUS: Q-H-[FYW]-S-x(4)-P-G.

NAME: Insulin family signature.
CONSENSUS: C-C-{P}-x(2)-C-[STDNEKPI]-x(3)-[LIVMFS]-x(3)-C.

NAME: Natriuretic peptides signature.
CONSENSUS: C-F-G-x(3)-D-R-I-x(3)-S-x(2)-G-C.

NAME: Neurohypophysial hormones signature.
CONSENSUS: C-[LIFY](2)-x-N-[CS]-P-x-G.

NAME: Neuromedin U signature.
CONSENSUS: F-[LIVMF]-F-R-P-R-N.

NAME: Endogenous opioids neuropeptides precursors signature.
CONSENSUS: C-x(3)-C-x(2)-C-x(2)-[KRH]-x(6,7)-[LIF]-[DN]-x(3)-C-x-[LIVM]-[EQ]-C-
CONSENSUS: [EQ]-x(8)-W-x(2)-C.

NAME: Pancreatic hormone family signature.
CONSENSUS: [FY]-x(3)-[LIVM]-x(2)-Y-x(3)-[LIVMFY]-x-R-x-R-[YF].

NAME: Parathyroid hormone family signature.
CONSENSUS: V-S-E-x-Q-x(2)-H-x(2)-G.

NAME: Pyrokinins signature.
CONSENSUS: F-[GSTV]-P-R-L-[G>].

NAME: Somatotropin, prolactin and related hormones signature 1.
CONSENSUS: C-x-[ST]-x(2)-[LIVMFY]-x-[LIVMSTA]-P-x(5)-[TALIV]-x(7)-[LIVMFY]-x(6)-
CONSENSUS: [LIVMFY]-x(2)-[STA]-W.

NAME: Somatotropin, prolactin and related hormones signature 2.
CONSENSUS: C-[LIVMFY]-x(2)-D-[LIVMFYSTA]-x(5)-[LIVMFY]-x(2)-[LIVMFYT]-x(2)-C.

NAME: Tachykinin family signature.
CONSENSUS: F-[IVFY]-G-[LM]-M-[G>].

NAME: Thymosin beta-4 family signature.
CONSENSUS: K-L-K-K-T-E-T-Q-E-K-N.

NAME: Urotensin II signature.
CONSENSUS: C-F-W-K-Y-C.

NAME: Cecropin family signature.
CONSENSUS: W-x(0,2)-[KDN]-x(2)-K-[KRE]-[LI]-E-[RKN].

NAME: Mammalian defensins signature.
CONSENSUS: C-x-C-x(3,5)-C-x(7)-G-x-C-x(9)-C-C.

NAME: Arthropod defensins signature.
CONSENSUS: C-x(2,3)-[HN]-C-x(3,4)-[GR]-x(2)-G-G-x-C-x(4,7)-C-x-C.

NAME: Cathelicidins signature 1.
CONSENSUS: Y-x-[ED]-x-V-x-[RQ]-A-[LIVMA]-[DQG]-x-[LIVMFY]-N-[EQ].

NAME: Cathelicidins signature 2.
CONSENSUS: F-x-[LIVM]-K-E-T-x-C-x(10)-C-x-F-[KR]-[KE].

NAME: Endothelin family signature.
CONSENSUS: C-x-C-x(4)-D-x(2)-C-x(2)-[FY]-C.

NAME: Plant thionins signature.
CONSENSUS: C-C-x(5)-R-x(2)-[FY]-x(2)-C.

NAME: Gamma-thionins family signature.
CONSENSUS: [KR]-x-C-x(3)-[SV]-x(2)-[FYWH]-x-[GF]-x-C-x(5)-C-x(3)-C.

NAME: Snake toxins signature.
CONSENSUS: G-C-x(1,3)-C-P-x(8,10)-C-C-x(2)-[PDEN].

NAME: Myotoxins signature.
CONSENSUS: K-x-C-H-x-K-x(2)-H-C-x(2)-K-x(3)-C-x(8)-K-x(2)-C-x(2)-[RK]-x-K-C-C-K-K.

NAME: Scorpion short toxins signature.
CONSENSUS: C-x(3)-C-x(6,9)-[GAS]-K-C-[IMQT]-x(3)-C-x-C.

NAME: Heat-stable enterotoxins signature.
CONSENSUS: C-C-x(2)-C-C-x-P-A-C-x-G-C.

NAME: Aerolysin type toxins signature.
CONSENSUS: [KT]-x(2)-N-W-x(2)-T-[DN]-T.

NAME: Shiga/ricin ribosomal inactivating toxins active site signature.
CONSENSUS: [LIVMA]-x-[LIVMSTA](2)-x-E-[SAGV]-[STAL]-R-[FY]-[RKNQS]-x-[LIVM]-[EQS]-
CONSENSUS: x(2)-[LIVMF].

NAME: Channel forming colicins signature.
CONSENSUS: T-x(2)-W-x-P-[LIVMFY](3)-x(2)-E.

NAME: Hok/gef family cell toxic proteins signature.
CONSENSUS: [LIVMA](4)-C-[LIVMFA]-T-[LIVMA](2)-x(4)-[LIVM]-x-[RG]-x(2)-L-[CY].

NAME: Staphylococcal enterotoxin/Streptococcal pyrogenic exotoxin signature 1.
CONSENSUS: Y-G-G-[LIV]-T-x(4)-N.

NAME: Staphyloccocal enterotoxin/Streptococcal pyrogenic exotoxin signature 2.
CONSENSUS: K-x(2)-[LIV]-x(4)-[LIV]-D-x(3)-R-x(2)-L-x(5)-[LIV]-Y.

NAME: Thiol-activated cytolysins signature.
CONSENSUS: [RK]-E-C-T-G-L-x-W-E-W-W-[RK].

NAME: Membrane attack complex components / perforin signature.
CONSENSUS: Y-x(6)-[FY]-G-T-H-[FY].

NAME: Pancreatic trypsin inhibitor (Kunitz) family signature.
CONSENSUS: F-x(3)-G-C-x(6)-[FY]-x(5)-C.

NAME: Bowman-Birk serine protease inhibitors family signature.
CONSENSUS: C-x(5,6)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDKS]-[DEKRHSTA]-C.

NAME: Kazal serine protease inhibitors family signature.
CONSENSUS: C-x(7)-C-x(6)-Y-x(3)-C-x(2,3)-C.

NAME: Soybean trypsin inhibitor (Kunitz) protease inhibitors family signature.
CONSENSUS: [LIVM]-x-D-x-[EDNTY]-[DG]-[RKHDENQ]-x-[LIVM]-x(5)-Y-x-[LIVM].

NAME: Serpins signature.
CONSENSUS: [LIVMFY]-x-[LIVMFYAC]-[DNQ]-[RKHQS]-[PST]-F-[LIVMFY]-[LIVMFYC]-x-
CONSENSUS: [LIVMFAH].

NAME: Potato inhibitor I family signature.
CONSENSUS: [FYW]-P-[EQH]-[LIV](2)-G-x(2)-[STAGV]-x(2)-A.

NAME: Squash family of serine protease inhibitors signature.
CONSENSUS: C-P-x(5)-C-x(2)-D-x-D-C-x(3)-C-x-C.

NAME: Streptomyces subtilisin-type inhibitors signature.
CONSENSUS: C-x-P-x(2,3)-G-x-H-P-x(4)-A-C-[ATD]-x-L.

NAME: Cysteine proteases inhibitors signature.
CONSENSUS: [GSTEQKRV]-Q-[LIVT]-[VAF]-[SAGQ]-G-x-[LIVMNK]-x(2)-[LIVMFY]-x-[LIVMFYA]-
CONSENSUS: [DENQKRHSIV].

NAME: Tissue inhibitors of metalloproteinases signature.
CONSENSUS: C-x-C-x-P-x-H-P-Q-x-A-F-C.

NAME: Cereal trypsin/alpha-amylase inhibitors family signature.
CONSENSUS: C-x(4)-[SAGD]-x(4)-[SPAL]-[LF]-x(2)-C-[RH]-x-[LIVMFY](2)-x(3,4)-C.

NAME: Alpha-2-macroglobulin family thiolester region signature.
CONSENSUS: [PG]-x-[GS]-C-[GA]-E-[EQ]-x-[LIVM].

NAME: Disintegrins signature.
CONSENSUS: C-x(2)-G-x-C-C-x-[NQRS]-C-x-[FM]-x(6)-C-[RK].

NAME: Lambdoid phages regulatory protein CIII signature.
CONSENSUS: E-S-x-L-x-R-x(2)-[KR]-x-L-x(4)-[KR](2)-x(2)-[DE]-x-L.

NAME: Chaperonins cpn60 signature.
CONSENSUS: A-[AS]-x-[DEQ]-E-x(4)-G-G-[GA].

NAME: Chaperonins cpn10 signature.
CONSENSUS: [LIVMFY]-x-P-[ILT]-x-[DEN]-[KR]-[LIVMFA](3)-[KREQ]-x(8,9)-[SG]-x-

CONSENSUS: [LIVMFY](3).

NAME: Chaperonins TCP-1 signature 1.
CONSENSUS: [RKEL]-[ST]-x-[LMFY]-G-P-x-[GSA]-x-x-K-[LIVMF](2).

NAME: Chaperonins TCP-1 signature 2.
CONSENSUS: [LIVM]-[TS]-[NK]-D-[GA]-[AVNHK]-[TAV]-[LIVM](2)-x(2)-[LIVM]-x-[LIVM]-x-
CONSENSUS: [SNH]-[PQH].

NAME: Chaperonins TCP-1 signature 3.
CONSENSUS: Q-[DEK]-x-x-[LIVMGTA]-[GA]-D-G-T.

NAME: Heat shock hsp20 proteins family profile.

NAME: Heat shock hsp70 proteins family signature 1.
CONSENSUS: [IV]-D-L-G-T-[ST]-x-[SC].

NAME: Heat shock hsp70 proteins family signature 2.
CONSENSUS: [LIVMF]-[LIVMFY]-[DN]-[LIVMFS]-G-[GSH]-[GS]-[AST]-x(3)-[ST]-[LIVM]-
CONSENSUS: [LIVMFC].

NAME: Heat shock hsp70 proteins family signature 3.
CONSENSUS: [LIVMY]-x-[LIVMF]-x-G-G-x-[ST]-x-[LIVM]-P-x-[LIVM]-x-[DEQKRSTA].

NAME: Heat shock hsp90 proteins family signature.
CONSENSUS: Y-x-[NQH]-K-[DE]-[IVA]-F-L-R-[ED].

NAME: Chaperonins clpA/B signature 1.
CONSENSUS: D-[AI]-[SGA]-N-[LIVMF](2)-K-[PT]-x-L-x(2)-G.

NAME: Chaperonins clpA/B signature 2.
CONSENSUS: R-[LIVMFY]-D-x-S-E-[LIVMFY]-x-E-[KRQ]-x-[STA]-x-[STA]-[KR]-[LIVM]-x-G-
CONSENSUS: [STA].

NAME: Nt-dnaJ domain signature.
CONSENSUS: [FY]-x(2)-[LIVMA]-x(3)-[FYWHNT]-[DENQSA]-x-L-x-[DN]-x(3)-[KR]-x(2)-[FYI].

NAME: dnaJ domain profile.

NAME: CXXCXGXG dnaJ domain signature.
CONSENSUS: C-[DEGSTHKR]-x-C-x-G-x-[GK]-[AGSDM]-x(2)-[GSNKR]-x(4,6)-C-x(2,3)-C-x-G-x-G.

NAME: grpE protein signature.
CONSENSUS: [FL]-[DN]-[PHEA]-x(2)-[HM]-x-A-[LIVMTN]-x(16,20)-G-[FY]-x(3)-[DEG]-x(2)-
CONSENSUS: [LIVM]-[RI]-x-[SA]-x-V-x-[IV].

NAME: Bacterial type II secretion system protein C signature.
CONSENSUS: P-x(6)-F-x(4)-L-x(3)-D-[LIVM]-A-[LIVM]-x-[LIVM]-N-x-[LIVM]-x-L.

NAME: Bacterial type II secretion system protein D signature.
CONSENSUS: [GR]-[DEQKG]-[STVM]-[LIVMA](3)-[GA]-G-[LIVMFY]-x(11)-[LIVM]-P-
CONSENSUS: [LIVMFYWGS]-[LIVMF]-[GSAE]-x-[LIVM]-P-[LIVMFYW](2)-x(2)-[LV]-F.

NAME: Bacterial type II secretion system protein E signature.
CONSENSUS: [LIVM]-R-x(2)-P-D-x-[LIVM](3)-G-E-[LIVM]-R-D.

NAME: Bacterial type II secretion system protein F signature.
CONSENSUS: [KRQ]-[LIVMA]-x(2)-[SAIV]-[LIVM]-x-[TY]-P-x(2)-[LIVM]-x(3)-[STAGV]-x(6)-
CONSENSUS: [LMY]-x(3)-[LIVMF](2)-P.

NAME: Bacterial type II secretion system protein N signature.
CONSENSUS: G-T-L-W-x-G-x(11)-L-x(4)-W.

NAME: Bacterial export FHIPEP family signature.
CONSENSUS: R-[LIVM]-[GSA]-E-V-[GSA]-A-R-F-[STV]-L-D-[GSA]-M-P-G-K-Q-M-[GSA]-I-D-
CONSENSUS: [GSA]-D.

NAME: Protein secA signatures.
CONSENSUS: [IV]-x-[IV]-[SA]-T-[NQ]-M-A-G-R-G-x-D-I-x-L.

NAME: Protein secY signature 1.
CONSENSUS: [GST]-[LIVMF](2)-x-[LIVM]-G-[LIVM]-x-P-[LIVMFY](2)-x-[AS]-[GSTQ]-
CONSENSUS: [LIVMFAT](3)-Q-[LIVMFA](2).

NAME: Protein secY signature 2.
CONSENSUS: [LIVMFYW](2)-x-[DE]-x-[LIVMF]-[STN]-x(2)-G-[LIVMF]-[GST]-[NST]-G-x-[GST]-
CONSENSUS: [LIVMF](3).

NAME: Protein secE/sec61-gamma signature.
CONSENSUS: [LIVMFY]-x(2)-[DENQGA]-x(4)-[LIVMTA]-x-[KRV]-x(2)-[KW]-P-x(3)-[SEQ]-x(7)-
CONSENSUS: [LIVT]-[LIVGA]-[LIVFGAST].

NAME: Gram-negative pili assembly chaperone signature.
CONSENSUS: [LIVMFY]-[APN]-x-[DNS]-[KREQ]-E-[STR]-[LIVMAR]-x-[FYWT]-x-[NC]-[LIVM]-
CONSENSUS: x(2)-[LIVM]-P-[PAS].

NAME: Fimbrial biogenesis outer membrane usher protein signature.
CONSENSUS: [VL]-[PASQ]-[PAS]-G-[PAD]-[FY]-x-[LI]-[DNQSTAP]-[DNH]-[LIVMFY].

NAME: SRP54-type proteins GTP-binding domain signature.
CONSENSUS: P-[LIVM]-x-[FYL]-[LIVMAT]-[GS]-x-[GS]-[EQ]-x(4)-[LIVMF].

NAME: Cytochrome c oxidase assembly factor COX10/ctaB/cyoE signature.
CONSENSUS: [ED]-x-D-x(2)-M-x-R-T-x(2)-R-x(4)-G.

NAME: Cyclin-dependent kinases regulatory subunits signature 1.
CONSENSUS: Y-S-x-[KR]-Y-x-[DE](2)-x-[FY]-E-Y-R-H-V-x-[LV]-[PT]-[KRP].

NAME: Cyclin-dependent kinases regulatory subunits signature 2.
CONSENSUS: H-x-P-E-x-H-[IV]-L-L-F-[KR].

NAME: Pentaxin family signature.
CONSENSUS: H-x-C-x-[ST]-W-x-[ST].

NAME: Immunoglobulins and major histocompatibility complex proteins signature.
CONSENSUS: [FY]-x-C-x-[VA]-x-H.

NAME: Prion protein signature 1.
CONSENSUS: A-G-A-A-A-A-G-A-V-V-G-G-L-G-G-Y.

NAME: Prion protein signature 2.
CONSENSUS: E-x-[ED]-x-K-[LIVM](2)-x-[KR]-[LIVM](2)-x-[QE]-M-C-x(2)-Q-Y.

NAME: Cyclins signature.
CONSENSUS: R-x(2)-[LIVMSA]-x(2)-[FYWS]-[LIVM]-x(8)-[LIVMFC]-x(4)-[LIVMFYA]-x(2)-
CONSENSUS: [STAGC]-[LIVMFYQ]-x-[LIVMFYC]-[LIVMFY]-D-[RKH]-[LIVMFYW].

NAME: Proliferating cell nuclear antigen signature 1.
CONSENSUS: [GA]-[LIVMF]-x-[LIVMA]-x-[SAV]-[LIVM]-D-x-[NSAE]-[HKR]-[VI]-x-[LY]-
CONSENSUS: [VGA]-x-[LIVM]-x-[LIVM]-x(4)-F.

NAME: Proliferating cell nuclear antigen signature 2.
CONSENSUS: [RKA]-C-[DE]-[RH]-x(3)-[LIVMF]-x(3)-[LIVM]-x-[SGAN]-[LIVMF]-x-K-
CONSENSUS: [LIVMF](2).

NAME: Actin-depolymerizing proteins signature.
CONSENSUS: P-[DE]-x-[SA]-x-[LIVMT]-[KR]-x-[KR]-M-[LIVM]-[YA]-[STA](3)-x(3)-[LIVMF]-
CONSENSUS: [KR].

NAME: BCL2-like apoptosis inhibitors (spans part of BH3, BH1 and BH2).

NAME: Apoptosis regulator, Bcl-2 family BH1 domain signature.
CONSENSUS: [LVME]-[FT]-x-[GSD]-[GL]-x(1,2)-[NS]-[YW]-G-R-[LIV]-[LIVC]-[GAT]-
CONSENSUS: [LIVMF](2)-x-F-[GSAE]-[GSARY].

NAME: Apoptosis regulator, Bcl-2 family BH2 domain signature.
CONSENSUS: W-[LIM]-x(3)-[GR]-G-[WQ]-[DENSAV]-x-[FLGA]-[LIVFTC].

NAME: Apoptosis regulator, Bcl-2 family BH3 domain signature.
CONSENSUS: [LIVAT]-x(3)-L-[KARQ]-x-[IVAL]-G-D-[DESG]-[LIMFV]-[DENSHQ]-[LVSHRQ]-
CONSENSUS: [NSR].

NAME: Apoptosis regulator, Bcl-2 family BH4 domain signature.
CONSENSUS: [DS]-[NT]-R-[AE]-[LI]-V-x-[KD]-[FY]-[LIV]-[GHS]-Y-K-L-[SR]-Q-[RK]-G-
CONSENSUS: [HY]-x-[CW].

NAME: Apoptosis regulator, Bcl-2 family BH4 domain profile.

NAME: Arrestins signature.
CONSENSUS: [FY]-R-Y-G-x-[DE](2)-x-[DE]-[LIVM](2)-G-[LIVM]-x-F-x-[RK]-[DEQ]-[LIVM].

NAME: AAA-protein family signature.
CONSENSUS: [LIVMT]-x-[LIVMT]-[LIVMF]-x-[GATMC]-[ST]-[NS]-x(4)-[LIVM]-D-x-A-[LIFA]-
CONSENSUS: x-R.

NAME: Ubiquitin domain signature.
CONSENSUS: K-x(2)-[LIVM]-x-[DESAK]-x(3)-[LIVM]-[PA]-x(3)-Q-x-[LIVM]-[LIVMC]-
CONSENSUS: [LIVMFY]-x-G-x(4)-[DE].

NAME: Ubiquitin domain profile.

NAME: ADP-ribosylation factors family signature.
CONSENSUS: [HRQT]-x-[FYWI]-x-[LIVM]-x(4)-A-x(2)-G-x(2)-[LIVM]-x(2)-[GSA]-[LIVMF]-x-
CONSENSUS: [WK]-[LIVM].

NAME: GTP-binding nuclear protein ran signature.
CONSENSUS: D-T-A-G-Q-E-K-[LF]-G-G-L-R-[DE]-G-Y-Y.

NAME: SAR1 family signature.
CONSENSUS: R-x-[LIVM]-E-V-F-M-C-S-[LIVM](2)-x-[KRQ]-x-G-Y-x-E-[AG]-[FI]-x-W-[LIVM]-
CONSENSUS: x-Q-Y.

NAME: Band 7 protein family signature.
CONSENSUS: R-x(2)-[LIV]-[SAN]-x(6)-[LIV]-D-x(2)-T-x(2)-W-G-[LIV]-[KRH]-[LIV]-x-
CONSENSUS: [KR]-[LIV]-E-[LIV]-[KR].

NAME: Trp-Asp (WD) repeats signature.
CONSENSUS: [LIVMSTAC]-[LIVMFYWSTAGC]-[LIMSTAG]-[LIVMSTAGC]-x(2)-[DN]-x(2)-
CONSENSUS: [LIVMWSTAC]-x-[LIVMFSTAG]-W-[DEN]-[LIVMFSTAGCN].

NAME: G-protein gamma subunit profile.

NAME: Ras GTPase-activating proteins signature.
CONSENSUS: [GSN]-x-[LIVMF]-[FY]-[LIVMFY]-R-[LIVMFY](2)-[GACN]-P-[AV]-[LIV](2)-
CONSENSUS: [SGAN]-P.

NAME: Ras GTPase-activating proteins profile.

NAME: Guanine-nucleotide dissociation stimulators CDC24 family signature.
CONSENSUS: L-x(2)-[LIVMFYW]-L-x(2)-P-[LIVM]-x(2)-[LIVM]-x-[KRS]-x(2)-L-x-[LIVM]-x-
CONSENSUS: [DEQ]-[LIVM]-x(3)-[ST].

NAME: Guanine-nucleotide dissociation stimulators CDC25 family signature.
CONSENSUS: [GAP]-[CT]-V-P-[FY]-x(4)-[LIVMFY]-x-[DN]-[LIVM].

NAME: MARCKS family signature 1.
CONSENSUS: G-Q-E-N-G-H-V-[KR].

NAME: MARCKS family phosphorylation site domain.
CONSENSUS: E-T-P-K(5)-x(0,1)-F-S-F-K-K-x-F-K-L-S-G-x-S-F-K-[KR]-[NS]-[KR]-K-E.

NAME: Stathmin family signature 1.
CONSENSUS: P-[KQ]-[KR](2)-[DE]-x-S-L-[EG]-E.

NAME: Stathmin family signature 2.
CONSENSUS: A-E-K-R-E-H-E-[KR]-E-V.

NAME: GTP-binding elongation factors signature.
CONSENSUS: D-[KRSTGANQFYW]-x(3)-E-[KRAQ]-x-[RKQD]-[GC]-[IVMK]-[ST]-[IV]-x(2)-
CONSENSUS: [GSTACKRNQ].

NAME: Elongation factor 1 beta/beta'/delta chain signature 1.
CONSENSUS: [DE]-[DEG]-[DE](2)-[LIVMF]-D-L-F-G.

NAME: Elongation factor 1 beta/beta'/delta chain signature 2.
CONSENSUS: V-Q-S-x-D-[LIVM]-x-A-[FWM]-[NQ]-K-[LIVM].

NAME: Elongation factor 1 gamma chain profile.

NAME: Elongation factor Ts signature 1.
CONSENSUS: L-R-x(2)-T-[GDQ]-x-[GS]-[LIVMF]-x(0,1)-[DENKAC]-x-K-[KRNEQS]-[AV]-L.

NAME: Elongation factor Ts signature 2.
CONSENSUS: E-[LIVM]-N-[SCV]-[QE]-T-D-F-V-[SA]-[KRN].

NAME: Elongation factor P signature.
CONSENSUS: K-x-A-x(4)-G-x(2)-[LIV]-x-V-P-x(2)-[LIV]-x(2)-G.

NAME: Eukaryotic initiation factor 1A signature.
CONSENSUS: [IM]-x-G-x-[GS]-[KRH]-x(4)-[CL]-x-D-G-x(2)-R-x(2)-[RH]-I-x-G.

NAME: Eukaryotic initiation factor 4E signature.
CONSENSUS: [DE]-[IFY]-x(2)-F-[KR]-x(2)-[LIVM]-x-P-x-W-E-[DV]-x(5)-G-G-[KR]-W.

NAME: Eukaryotic initiation factor 5A hypusine signature.
CONSENSUS: [PT]-G-K-H-G-x-A-K.

NAME: Initiation factor 2 signature.
CONSENSUS: G-x-[LIVM]-x(2)-L-[KR]-[KRHNS]-x-K-x(5)-[LIVM]-x(2)-G-x-[DEN]-C-G.

NAME: Initiation factor 3 signature.
CONSENSUS: [KR]-[LIVM](2)-[DN]-[FY]-[GSN]-[KR]-[LIVMFYS]-x-[FY]-[DEQT]-x(2)-[KR].

NAME: Translation initiation factor SUI1 signature.
CONSENSUS: [LIVM]-[EQ]-[LIVM]-Q-G-[DEN]-[KHQ]-[KRV].

NAME: Prokaryotic-type class I peptide chain release factors signature.
CONSENSUS: [AR]-[STA]-x-G-x-G-G-Q-[HNGCS]-V-N-x(3)-[ST]-A-[IV].

NAME: Transcription termination factor nusG signature.
CONSENSUS: [LIVM]-F-G-[KRW]-x-T-P-[IV]-x-[LIVM].

NAME: Calponin family repeat.
CONSENSUS: [LIVM]-x-[LS]-Q-[MAS]-G-[STY]-[NT]-[KRQ]-x(2)-[STN]-Q-x-G-x(3,4)-G.

NAME: CAP protein signature 1.
CONSENSUS: [LIVM](2)-x-R-L-[DE]-x(4)-R-L-E.

NAME: CAP protein signature 2.
CONSENSUS: D-[LIVMFY]-x-E-x-[PA]-x-P-E-Q-[LIVMFY]-K.

NAME: Calreticulin family signature 1.
CONSENSUS: [KRHN]-x-[DEQN]-[DEQNK]-x(3)-C-G-G-[AG]-[FY]-[LIVM]-[KN]-[LIVMFY](2).

NAME: Calreticulin family signature 2.
CONSENSUS: [LIVM](2)-F-G-P-D-x-C-[AG].

NAME: Calreticulin family repeated motif signature.
CONSENSUS: [IV]-x-D-x-[DENST]-x(2)-K-P-[DEH]-D-W-[DEN].

NAME: Calsequestrin signature 1.
CONSENSUS: [EQ]-[DE]-G-L-[DN]-F-P-x-Y-D-G-x-D-R-V.

NAME: Calsequestrin signature 2.
CONSENSUS: [DE]-L-E-D-W-[LIVM]-E-D-V-L-x-G-x-[LIVM]-N-T-E-D-D-D.

NAME: S-100/ICaBP type calcium binding protein signature.
CONSENSUS: [LIVMFYW](2)-x(2)-[LK]-D-x(3)-[DN]-x(3)-[DNSG]-[FY]-x-[ES]-[FYVC]-x(2)-
CONSENSUS: [LIVMFS]-[LIVMF].

NAME: Hemolysin-type calcium-binding region signature.
CONSENSUS: D-x-[LI]-x(4)-G-x-D-x-[LI]-x-G-G-x(3)-D.

NAME: HlyD family secretion proteins signature.
CONSENSUS: [LIVM]-x(2)-G-[LM]-x(3)-[STGAV]-x-[LIVMT]-x-[LIVMT]-[GE]-x-[KR]-x-
CONSENSUS: [LIVMFYW](2)-x-[LIVMFYW](3).

NAME: P-II protein urydylation site.
CONSENSUS: Y-[KR]-G-[AS]-[AE]-Y.

NAME: P-II protein C-terminal region signature.
CONSENSUS: [ST]-x(3)-G-[DY]-G-[KR]-[IV]-[FW]-[LIVM]-x(2)-[LIVM].

NAME: 14-3-3 proteins signature 1.
CONSENSUS: R-N-L-[LIV]-S-[VG]-[GA]-Y-[KN]-N-[IVA].

NAME: 14-3-3 proteins signature 2.
CONSENSUS: Y-K-[DE]-S-T-L-I-[IM]-Q-L-[LF]-[RHC]-D-N-[LF]-T-[LS]-W-[TAN]-[SAD].

NAME: ATP1G1 / PLM / MAT8 family signature.
CONSENSUS: [DNS]-x-F-x-Y-D-x(2)-[ST]-[LIVM]-[RQ]-x(2)-G.

NAME: BTG1 family signature 1.
CONSENSUS: Y-x(2)-[HP]-W-[FY]-[AP]-E-x-P-x-K-G-x-[GA]-[FY]-R-C-[IV]-[RH]-[IV].

NAME: BTG1 family signature 2.
CONSENSUS: [LV]-P-x-[DE]-[LM]-[ST]-[LIVM]-W-[IV]-D-P-x-E-V-[SC]-x-[RQ]-x-G-E.

NAME: Cullin family signature.
CONSENSUS: [LIV]-K-x(2)-[LIV]-x(2)-L-I-[DEQ]-[KRHNQ]-x-Y-[LIVM]-x-R-x(6,7)-[FY]-x-
CONSENSUS: Y-x-[SA]>.

NAME: Cullin family profile.

NAME: Enhancer of rudimentary signature.
CONSENSUS: Y-D-I-[SA]-x-L-[FY]-x-F-[IV]-D-x(3)-D-[LIV]-S.

NAME: G10 protein signature 1.
CONSENSUS: L-C-C-x-[KR]-C-x(4)-[DE]-x-N-x(4)-C-x-C-R-V-P.

NAME: G10 protein signature 2.
CONSENSUS: C-x-H-C-G-C-[KRH]-G-C-[SA].

NAME: Glucokinase regulatory protein family signature.
CONSENSUS: G-[PA]-E-x-[LIV]-[STA]-G-S-[ST]-R-[LIVM]-K-[STGA](3)-x(2)-K.

NAME: GTP1/OBG family signature.
CONSENSUS: D-[LIVM]-P-G-[LIVM](2)-[DEY]-[GN]-A-x(2)-G-x-G.

NAME: HIT family signature.
CONSENSUS: [NQA]-x(4)-[GAV]-x-[QF]-x-[LIVM]-x-H-[LIVMFYT]-H-[LIVMFT]-H-[LIVMF](2)-
CONSENSUS: [PSGA].

NAME: Caseins alpha/beta signature.
CONSENSUS: C-L-[LV]-A-x-A-[LVF]-A.

NAME: Clathrin adaptor complexes medium chain signature 1.
CONSENSUS: [IVT]-[GSP]-W-R-x(2,3)-[GAD]-x(2)-[HY]-x(2)-N-x-[LIVMAFY](3)-D-[LIVM]-
CONSENSUS: [LIVMT]-E.

NAME: Clathrin adaptor complexes medium chain signature 2.
CONSENSUS: [LIV]-x-F-I-P-P-x-G-x-[LIVMFY]-x-L-x(2)-Y.

NAME: Clathrin adaptor complexes small chain signature.
CONSENSUS: [LIVM](2)-Y-[KR]-x(4)-L-Y-F.

NAME: Ependymins signature 1.
CONSENSUS: F-E-E-G-x-[LIVMF]-Y-[ED]-I-D-x(2)-N-[QE]-S-C-[RKH](2).

NAME: Ependymins signature 2.
CONSENSUS: [QE]-[LIVMA]-F-x(2)-P-[STA]-[FY]-C-[DE]-[GA]-[LIVM]-x(2)-[DE](2).

NAME: Syntaxin / epimorphin family signature.
CONSENSUS: [RQ]-x(3)-[LIVMA]-x(2)-[LIVM]-[ESH]-x(2)-[LIVMT]-x-[DEVM]-[LIVM]-x(2)-
CONSENSUS: [LIVM]-[FS]-x(2)-[LIVM]-x(3)-[LIVT]-x(2)-Q-[GADEQ]-x(2)-[LIVM]-[DNQT]-x-
CONSENSUS: [LIVMF]-[DESV]-x(2)-[LIVM].

NAME: Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 1.
CONSENSUS: [GDER]-H-[FYWH]-T-Q-[LIVM](2)-W-x(2)-[STN].

NAME: Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 2.
CONSENSUS: [LIVMFYH]-[LIVMFY]-x-C-[NQRHS]-Y-x-[PARH]-x-[GL]-N-[LIVMFYWDN].

NAME: Fetuin family signature 1.
CONSENSUS: C-x(56)-C-x(10)-C-x(13)-C-x(17,18)-C-x(13)-C-x(2)-C-x(58)-C-x(10,11)-
CONSENSUS: C-x(10,12)-C-x(16,22)-C.

NAME: Fetuin family signature 2.
CONSENSUS: L-E-T-x-C-H-x-L-D-P-T-P.

NAME: Legume lectins beta-chain signature.
CONSENSUS: [LIV]-[STAG]-V-[DEQV]-[FLI]-D-[ST].

NAME: Legume lectins alpha-chain signature.
CONSENSUS: [LIV]-x-[EDQ]-[FYWKR]-V-x-[LIV]-G-[LF]-[ST].

NAME: Vertebrate galactoside-binding lectin signature.
CONSENSUS: W-[GEK]-x-[EQ]-x-[KRE]-x(3,6)-[PCTF]-[LIVMF]-[NQEGSKV]-x-[GH]-x(3)-
CONSENSUS: [DENKHS]-[LIVMFC].

NAME: Lysosome-associated membrane glycoproteins duplicated domain signature.
CONSENSUS: [STA]-C-[LIVM]-[LIVMFYW]-A-x-[LIVMFYW]-x(3)-[LIVMFYW]-x(3)-Y.

NAME: LAMP glycoproteins transmembrane and cytoplasmic domain signature.
CONSENSUS: C-x(2)-D-x(3,4)-[LIVM](2)-P-[LIVM]-x-[LIVM]-G-x(2)-[LIVM]-x-G-[LIVM](2)-
CONSENSUS: x-[LIVM](4)-A-[FY]-x-[LIVM]-x(2)-[KR]-[RH]-x(1,2)-[STAG](2)-Y-[EQ].

NAME: Glycophorin A signature.
CONSENSUS: I-I-x-[GAC]-V-M-A-G-[LIVM](2).

NAME: PMP-22 / EMP / MP20 family signature 1.
CONSENSUS: [LIVMF](4)-[SA]-T-x(2)-[DNKS]-x-W-x(9,13)-[LIV]-W-x(2)-C.

NAME: PMP-22 / EMP / MP20 family signature 2.
CONSENSUS: [RQ]-[AV]-x-M-[IV]-L-S-x-[LI]-x(4)-[GSA]-[LIVMF](3).

NAME: Oxysterol-binding protein family signature.
CONSENSUS: E-[KQ]-x-S-H-[HR]-P-P-x-[STACF]-A.

NAME: Yeast PIR proteins repeats signature.
CONSENSUS: S-Q-[IV]-[STGNH]-D-G-Q-[LIV]-Q-[AIV]-[STA].

NAME: Seminal vesicle protein I repeats signature.
CONSENSUS: [IVM]-x-G-Q-D-x-V-K-x(5)-[KN]-G-x(3)-[STLV].

NAME: Seminal vesicle protein II repeats signature.
CONSENSUS: [GSA]-Q-x-K-S-[FY]-x-Q-x-K-[SA].

NAME: Serum amyloid A proteins signature.
CONSENSUS: A-R-G-N-Y-[ED]-A-x-[QKR]-R-G-x-G-G-x-W-A.

NAME: Spermadhesins family signature 1.
CONSENSUS: C-G-x(2)-[LI]-x(4)-G-x-I-x(9)-C-x-W-T.

NAME: Spermadhesins family signature 2.
CONSENSUS: C-x-K-E-x-[LIVM]-E-[LIVM]-x-[DE]-x(3)-[GS]-x(5)-K-x-C.

NAME: Stress-induced proteins SRP1/TIP1 family signature.
CONSENSUS: P-W-Y-[ST](2)-R-L.

NAME: Glypicans signature.
CONSENSUS: C-x(2)-C-x-G-[LIVM]-x(4)-P-C-x(2)-[FY]-C-x(2)-[LIVM]-x(2)-G-C.

NAME: Syndecans signature.
CONSENSUS: [FY]-R-[IM]-[KR]-K(2)-D-E-G-S-Y.

NAME: Tissue factor signature.
CONSENSUS: W-K-x-K-C-x(2)-T-x-[DEN]-T-E-C-D-[LIVM]-T-D-E.

NAME: Translationally controlled tumor protein signature 1.
CONSENSUS: [IA]-G-[GAS]-N-[PA]-S-A-E-[GDE]-[PAGE]-x(0,1)-[DEG]-x-[DEN]-x(2)-[DE].

NAME: Translationally controlled tumor protein signature 2.
CONSENSUS: [FL]-[FY]-[IVT]-G-E-x-[MA]-x(2,5)-[DEN]-[GAS]-x-[LV]-[AV]-x(3)-[FY]-[KR]-
CONSENSUS: [DE].

NAME: Tub family signature 1.
CONSENSUS: F-[KHQ]-G-R-V-[ST]-x-A-S-V-K-N-F-Q.

NAME: Tub family signature 2.
CONSENSUS: A-F-[AG]-I-[SAC]-[LIVM]-[ST]-S-F-x-[GST]-K-x-A-C-E.

NAME: HCP repeats signature.
CONSENSUS: H-R-H-R-G-H-x(2)-[DE](7).

NAME: Bacterial ice-nucleation proteins octamer repeat.
CONSENSUS: A-G-Y-G-S-T-x-T.

NAME: Cell cycle proteins ftsW / rodA / spoVE signature.
CONSENSUS: [NV]-x(5)-[GTR]-[LIVMA]-x-P-[PTLIVM]-x-G-[LIVM]-x(3)-[LIVMFW](2)-S-[YSA]-
CONSENSUS: G-G-[STN]-[SA].

NAME: Enterobacterial virulence outer membrane protein signature 1.
CONSENSUS: G-[LIVMFY]-N-[LIVM]-K-Y-R-Y-E.

NAME: Enterobacterial virulence outer membrane protein signature 2.
CONSENSUS: [FYW]-x(2)-G-x-G-Y-[KR]-F>.

NAME: Hydrogenases expression/synthesis hypA family signature.
CONSENSUS: F-[CSA]-[FY]-[DE]-[LIVA](2)-x(3)-[ST]-[LIVM]-x(16)-C-x(2)-C-x(12,15)-
CONSENSUS: C-P-x-C.

NAME: Hydrogenases expression/synthesis hupF/hypC family signature.
CONSENSUS: <M-C-[LIV]-[GA]-[LIV]-P-x-[QKR]-[LIV].

NAME: Staphylocoagulase repeat signature.
CONSENSUS: A-R-P-x(3)-K-x-S-x-T-N-A-Y-N-V-T-T-x(2)-[DN]-G-x(3)-Y-G.

NAME: 11-S plant seed storage proteins signature.
CONSENSUS: N-G-x-[DE](2)-x-[LIVMF]-C-[ST]-x(11,12)-[PAG]-D.

NAME: Dehydrins signature 1.
CONSENSUS: S(5)-[DE]-x-[DE]-G-x(1,2)-G-x(0,1)-[KR](4).

NAME: Dehydrins signature 2.
CONSENSUS: [KR]-[LIM]-K-[DE]-K-[LIM]-P-G.

NAME: Germin family signature.
CONSENSUS: G-x(4)-H-x-H-P-x-A-x-E-[LIVM].

NAME: Oleosins signature.
CONSENSUS: [AG]-[ST]-x(2)-[AG]-x(2)-[LIVM]-[SAD]-T-P-[LIVMF](4)-F-S-P-[LIVM](3)-
CONSENSUS: P-A.

NAME: Small hydrophilic plant seed proteins signature.
CONSENSUS: G-[EQ]-T-V-V-P-G-G-T.

NAME: Pathogenesis-related proteins BetvI family signature.
CONSENSUS: G-x(2)-[LIVMF]-x(4)-E-x(2)-[CSTAEN]-x(8,9)-[GND]-G-[GS]-[CS]-x(2)-K-x(4)-
CONSENSUS: [FY].

NAME: Pollen proteins Ole e I family signature.
CONSENSUS: [EQ]-G-x-V-Y-C-D-T-C-R.

NAME: Thaumatin family signature.
CONSENSUS: G-x-[GF]-x-C-x-T-[GA]-D-C-x(1,2)-G-x(2,3)-C.

NAME: Mrp family signature.
CONSENSUS: W-x(2)-[LIVM]-D-[LIVMY](4)-D-x-P-P-G-T-[GS]-D.

NAME: Glucose inhibited division protein A family signature 1.
CONSENSUS: [GS]-P-x-Y-C-P-S-[LIVM]-E-x-K-[LIVM]-x-[KR]-F.

NAME: Glucose inhibited division protein A family signature 2.
CONSENSUS: A-G-Q-x-[NT]-G-x(2)-G-Y-x-E-[SAG](3)-[QS]-G-[LIVM](2)-A-G-[LIVMT]-N-A.

NAME: NOL1/NOP2/sun family signature.
CONSENSUS: [FV]-D-[KRA]-[LIVMA]-L-x-D-[AV]-P-C-[ST]-[GA].

NAME: PET112 family signature.
CONSENSUS: [DN]-x-[DN]-R-x(3)-P-L-[LIV]-E-[LIV]-x-[ST]-x-P.

NAME: Protein smpB signature.
CONSENSUS: [TA]-G-[LIVM]-x-L-x-G-x-E-[LIVM]-[KQ]-[SA]-[LIVM].

NAME: Hypothetical cof family signature 1.
CONSENSUS: [LIVFYAN]-[LIVMFA]-x(2)-D-[LIVMF]-[ND]-G-T-[LIV]-[LVY]-[STANLM].

NAME: Hypothetical cof family signature 2.
CONSENSUS: [LIVMFC]-G-D-[GSANQ]-x-N-D-x(3)-[LIMFY]-x(2)-[AV]-x(2)-[GSCP]-x(2)-
CONSENSUS: [LMP]-x(2)-[GAS].

NAME: RIO1/ZK632.3/MJ0444 family signature.
CONSENSUS: [LIVM]-V-H-[GA]-D-L-S-E-[FY]-N-x-[LIVM].

NAME: SUA5/yciO/yrdC family signature.
CONSENSUS: [LIVMTA](3)-[LIVMFYC]-[PG]-T-[DE]-[STA]-x-[FY]-[GA]-[LIVM]-[GS].

NAME: Uncharacterized protein family UPF0001 signature.
CONSENSUS: [FW]-H-[FM]-[IV]-G-x-[LIV]-Q-x-[NKR]-K-x(3)-[LIV].

NAME: Uncharacterized protein family UPF0003 signature.
CONSENSUS: G-x-V-x(2)-[LIV]-x(3)-[SA]-x(6)-D-x(3)-[LIVT](3)-P-N-x(2)-[LIVMF](2)-
CONSENSUS: x(5)-N.

NAME: Uncharacterized protein family UPF0004 signature.
CONSENSUS: [LIVM]-x-[LIVMT]-x(2)-G-C-x(3)-C-[STAN]-[FY]-C-x-[LIVM]-x(4)-G.

NAME: Uncharacterized protein family UPF0005 signature.
CONSENSUS: G-[LIVM](2)-[SA]-x(5,8)-G-x(2)-[LIVM]-G-P-x-L-x(4)-[SAG]-x(4,6)-
CONSENSUS: [LIVM](2)-x(2)-A-x(3)-T-A-[LIVM](2)-F.

NAME: Uncharacterized protein family UPF0006 signature 1.
CONSENSUS: [LIVMFY](2)-D-[STA]-H-x-H-[LIVMF]-[DN].

NAME: Uncharacterized protein family UPF0006 signature 2.
CONSENSUS: P-[LIVM]-x-[LIVM]-H-x-R-x-[TA]-x-[DE].

NAME: Uncharacterized protein family UPF0006 signature 3.
CONSENSUS: [LVSA]-[LIVA]-x(2)-[LIVM]-[PS]-x(3)-L-[LIVM]-[LIVMS]-E-T-D-x-P.

NAME: Uncharacterized protein family UPF0007 signature.
CONSENSUS: V-L-[IV]-H-D-[GA]-A-R.

NAME: Uncharacterized protein family UPF0011 signature.
CONSENSUS: S-D-A-G-x-P-x-[LIV]-[SN]-D-P-G.

NAME: Uncharacterized protein family UPF0012 signature.
CONSENSUS: [GTA]-x(2)-[IVT]-C-Y-D-[LIVM]-x-F-P-x(9)-G.

NAME: Uncharacterized protein family UPF0015 signature.
CONSENSUS: [DE]-[LIVMF](3)-R-T-[SG]-G-x(2)-R-x-S-x-[FY]-[LIVM](2)-W-Q.

NAME: Uncharacterized protein family UPF0016 signature.
CONSENSUS: E-[LIVM]-G-D-K-T-F-[LIVMF](2)-A.

NAME: Uncharacterized protein family UPF0017 signature.
CONSENSUS: D-x(8)-[GN]-[LFY]-x(4)-[DET]-[LY]-Y-x(3)-[ST]-x(7)-[IV]-x(2)-[PS]-x-
CONSENSUS: [LIVM]-x-[LIVM]-x(3)-[DN]-D.

NAME: Uncharacterized protein family UPF0019 signature.
CONSENSUS: L-P-V-[VT]-[NQL]-F-[AT]-A-G-G-[LIV]-A-T-P-A-D-A-A-[LM].

NAME: Uncharacterized protein family UPF0020 signature.
CONSENSUS: D-P-[LIVMF]-C-G-[ST]-G-x(3)-[LI]-E.

NAME: Uncharacterized protein family UPF0021 signature.
CONSENSUS: C-K-x(2)-F-x(4)-E-x(22,23)-S-G-G-K-D.

NAME: Uncharacterized protein family UPF0023 signature.
CONSENSUS: D-x-D-E-[LIV]-L-x(4)-V-F-x(3)-S-K-G.

NAME: Uncharacterized protein family UPF0024 signature.
CONSENSUS: G-x-K-D-[KR]-x-A-[LV]-T-x-Q-x-[LIVF]-[SGC].

NAME: Uncharacterized protein family UPF0025 signature.
CONSENSUS: D-V-[LIV]-x(2)-G-H-[ST]-H-x(12)-[LIVMF]-N-P-G.

NAME: Uncharacterized protein family UPF0027 signature.
CONSENSUS: Q-[LIVM]-x-N-x-A-x-[LIVM]-P-x-I-x(6)-[LIVM]-P-D-x-H-x-G-x-G-x(2)-[IV]-G.

NAME: Uncharacterized protein family UPF0028 signature.

CONSENSUS: [GA]-[GS]-G-[GA]-A-R-G-x-[SA]-H-x-G-x(9)-[IV]-x-[IV]-D-x(2)-[GA]-G-x-S-
CONSENSUS: x-G.

NAME: Uncharacterized protein family UPF0029 signature.
CONSENSUS: G-x(2)-[LIVM](2)-x(2)-[LIVM]-x(4)-[LIVM]-x(5)-[LIVM](2)-x-R-[FYW](2)-G-
CONSENSUS: G-x(2)-[LIVM]-G.

NAME: Uncharacterized protein family UPF0030 signature.
CONSENSUS: [GA]-L-I-[LIV]-P-G-G-E-S-T-[STA].

NAME: Uncharacterized protein family UPF0031 signature 1.
CONSENSUS: [SAV]-[IVW]-[LVA]-[LIV]-G-[PNS]-G-L-[GP]-x-[DENQT].

NAME: Uncharacterized protein family UPF0031 signature 2.
CONSENSUS: [GA]-G-x-G-D-[TV]-[LT]-[STA]-G-x-[LIVM].

NAME: Uncharacterized protein family UPF0032 signature.
CONSENSUS: Y-x(2)-F-[LIVMA](2)-x-L-x(4)-G-x(2)-F-[EQ]-[LIVMF]-P-[LIVM].

NAME: Uncharacterized protein family UPF0033 signature.
CONSENSUS: L-[DN]-x(2)-[TAG]-x(2)-C-P-x-P-x-[LIVM].

NAME: Uncharacterized protein family UPF0034 signature.
CONSENSUS: [LIVM]-[DNG]-[LIVM]-N-x-G-C-P-x(3)-[LIVMASQ]-x(5)-G-[SAC].

NAME: Uncharacterized protein family UPF0035 signature.
CONSENSUS: L-L-T-x-R-[SA]-x(3)-R-x(3)-G-x(3)-F-P-G-G.

NAME: Uncharacterized protein family UPF0036 signature.
CONSENSUS: H-x-S-G-H-[GA]-x(3)-[DE]-x(3)-[LM]-x(5)-P-x(3)-[LIVM]-P-x-H-G-[DE].

NAME: Uncharacterized protein family UPF0038 signature.
CONSENSUS: G-x-[LI]-x-R-x(2)-L-x(4)-F-x(8)-[LIV]-x(5)-P-x-[LIV].

NAME: Uncharacterized protein family UPF0044 signature.
CONSENSUS: L-[ST]-x(3)-K-x(3)-[KR]-[SGA]-x-[GA]-H-x-L-x-P-[LIV]-x(2)-[LIV]-[GA]-
CONSENSUS: x(2)-G.

NAME: Uncharacterized protein family UPF0047 signature.
CONSENSUS: S-X(2)-[LIV]-x-[LIV]-x(2)-G-x(4)-G-T-W-Q-x-[LIV].

NAME: Uncharacterized protein family UPF0054 signature.
CONSENSUS: H-[GS]-x-L-H-L-[LI]-G-[FYW]-D-H.

NAME: Uncharacterized protein family UPF0057 signature.
CONSENSUS: [LIV]-x-[STA]-[LIVF](3)-P-P-[LIVA]-[GA]-[IV]-x(4)-[GKN].

NAME: Hypothetical YER057c/yjjV family signature.
CONSENSUS: P-[AT]-R-[SA]-x-[LIVMY]-x(2)-[AK]-x-L-P-x(4)-[LIVM]-E.

NAME: Hypothetical hesB/yadR/yfhF family signature.
CONSENSUS: F-x-[LIVMFY]-x-N-[PG]-[NSK]-x(4)-C-x-C-[GS]-x-S-F.

NAME: Hypothetical yabO/yceC/sfhB family signature.
CONSENSUS: [NHY]-R-[LI]-D-x(2)-T-[ST]-G-[LIVMA]-[LIVMF](2)-[LIVMFG]-[SGAC].

We claim:

1. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3b16; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_41m15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4;; hfbr1_10e4; hfbr2_82g14;; hfbr1_10g14; hfbr2_82i17;; hfbr1_10; hfbr2_82i24;; hfbr1_10; hfbr2_82m16;; hfbr1_10; hfbr2_82m6;; hfbr1_10; hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; hmcf1_1a11; hmcf1_1c23; hmcf1_1e15; hmcf1_1g13; hhtes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17l17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18l7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21l16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23l11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2l19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2ol3; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06;

htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; hute1_17k7; hute1_18c12; hute1_18i19; hute1_18i4; hute1_18l1; hute1_19f19; hute1_19g19; hute1_19g22; hute1_19h17; hute1_19j11; hute1_1i2; hute1_20b19; hute1_20g21; hute1_20h13; hute1_20m11; hute1_20m24; hute1_21d15; hute1_22d2; hute1_22e12; hute1_22n2; hute1_22o2; hute1_23e13; hute1_23g11; hute1_24c19; hute1_24e11; hute1_24j6; hute1_2h3; their complements; and variants thereof.

2. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; ; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_41m15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4; hfbr1_10e4; hfbr2_82g14; hfbr1_10g14; hfbr2_82i17; hfbr1_10; hfbr2_82i24; hfbr1_10; hfbr2_82m16; hfbr1_10; hfbr2_82m6; hfbr1_10; their complements; and variants thereof.

3. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16f21; hfbr2_16k22; hfbr2_22f21; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23f2; ; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2c1; hfbr2_2c18; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2k19; hfbr2_3f16; hfbr2_3l2; hfbr2_62n10; hfbr2_64a11; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64o16; hfbr2_6a17; hfbr2_6i20; hfbr2_71o20;

hfbr2_72d13; hfbr2_72m16; hfbr2_72n12; hfbr2_78d13; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82m16; and hfbr1_10.

4. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; their complements; and variants thereof.

5. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfkd2_1j9; hfkd2_24e23; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_4b6; hfkd2_4c8; their complements; and variants thereof.

6. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; their complements; and variants thereof.

7. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hmcfl_1c23 hmcfl_1g13; their complements; and variants thereof.

8. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hhtes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17l17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18l7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21l16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23l11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2l19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2o13; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21;

htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06; htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; their complements; and variants thereof.

9. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: htes3_14g5; htes3_14p14; htes3_14p7; htes3_15a13; htes3_15g14; htes3_15h1; htes3_15j18; htes3_17f10; Htes3_18f3; htes3_19f19; htes3_19j17; htes3_20c21; htes3_21n23; htes3_22c23; htes3_22n13; Htes3_23n19; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2d15; htes3_2f14; htes3_2g7; htes3_2h15; htes3_2l19; htes3_2m20; htes3_2n9; htes3_30f4; htes3_35g6; htes3_35n24; htes3_35p17; htes3_4b4; htes3_4f17; htes3_4o19; htes3_50j4; htes3_50n23; htes3_50n06; htes3_6b21; htes3_6d16; htes3_72k11; htes3_7d17; htes3_7j8; Htes3_8g11; Htes3_8g5; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; their complements; and variants thereof.

10. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16g18; hfbr2_2k14; Htes3_35b4; htes3_35p22; htes3_7j3; htes3_7p10; hutel_20m11; their complements; and variants thereof.

11. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_2b5; htes3_15i5; htes3_18l7; htes3_1k11; Htes3_72k15; htes3_7b22; hutel_19g22; hutel_24j6; their complements; and variants thereof.

12. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_2d15; htes3_35e21; hutel_2h3; their complements; and variants thereof.

13. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_23l24; hfbr2_2i17; hfbr2_41m15; hfbr2_62f10; hfbr2_62l19; hfbr2_64j18;

hfkd2_24n20; hfkd2_24p5; hfkd2_4k14; htes3_1g13; htes3_21l16; htes3_23l11; htes3_26g22; htes3_4h6; htes3_72p16; hute1_19h17; hute1_20h13; hute1_24e11; their complements; and variants thereof.

14. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_3g8; hfbr2_62o17; hfbr2_6b24; hfbr2_78k24; hfkd2_24b15; hfkd2_3o17; hfkd2_46j20; htes3_17l17; htes3_17n18; htes3_27d1; htes3_2a17; htes3_35b5; htes3_35k16; htes3_35n12; htes3_35n9; hute1_20b19; hute1_20m24; hute1_23e13; their complements; and variants thereof.

15. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_23b10; hfbr2_3c18; hfbr2_64a15; hfbr2_6o17; hfbr2_72b18; hfbr2_72l12; hfbr2_82i24(hfbr1_10); htes3_14h21; Htes3_15j3; htes3_20m18; htes3_22g2; htes3_2m18; htes3_7p9; htes3_8m10; hute1_18l1; their complements; and variants thereof.

16. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_23b21; hfbr2_23n16; hfbr2_2c17; hfbr2_62b11; hfbr2_78c24; hfbr2_82e4 (hfbr1_10e4); hfbr2_82i17 (hfbr1_10); hfbr2_82m6 (hfbr1_10); hfkd2_46m4; htes3_15k11; htes3_1c1; hhtes3_1n3; htes3_20k2; htes3_21d4; htes3_23n19; htes3_4f5; htes3_6c11; htes3_8e24; hute1_20g21; hute1_22d2; hute1_22e12; their complements; and variants thereof.

17. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfbr2_16i12; hfbr2_16l12; hfbr2_22h13; hfbr2_2b17; hfbr2_2d17; hfbr2_64k24; hfbr2_82c20 (hfbr1_10c20); hfbr2_82e17 (hfbr1_10e17); hfbr2_82g14 (hfbr1_10g14); hfkd2_24a15; hfkd2_3i13; hfkd2_4m11; hmcf1_1a11; hmcf1_1e15; htes3_15c6; htes3_2o13; htes3_27k4; htes3_2h1; htes3_35k24; hute1_19f19; and hute1_24c19; their complements; and variants thereof.

18. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfkd2_46k19; hfkd2_47a4;

htes3_2e12; htes3_21j15; htes3_17n12; hutel_18i19; hutel_1i2; their complements; and variants thereof.

19. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hutel_17k7; hutel_18c12; hutel_18i19; hutel_18i4; hutel_18l1; hutel_19f19; hutel_19g19; hutel_19g22; hutel_19h17; hutel_19j11; hutel_1i2; hutel_20b19; hutel_20g21; hutel_20h13; hutel_20m11; hutel_20m24; hutel_21d15; hutel_22d2; hutel_22e12; hutel_22n2; hutel_22o2; hutel_23e13; hutel_23g11; hutel_24c19; hutel_24e11; hutel_24j6; hutel_2h3; their complements; and variants thereof.

20. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hutel_17k7; hutel_18c12; hutel_18i4; hutel_19g19; hutel_19j11; hutel_22n2; hutel_21d15; hutel_22o2; hutel_23g11; their complements; and variants thereof.

21. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; ; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_41m15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4;; hfbr1_10e4; hfbr2_82g14;; hfbr1_10g14; hfbr2_82i17;; hfbr1_10; hfbr2_82i24;; hfbr1_10; hfbr2_82m16;; hfbr1_10; hfbr2_82m6;; hfbr1_10; hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6;

hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; hhtes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17l17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18l7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21l16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23l11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2l19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2o13; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06; htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; hutel_17k7; hutel_18c12; hutel_18i19; hutel_18i4; hutel_18l1; hutel_19f19; hutel_19g19; hutel_19g22; hutel_19h17; hutel_19j11; hutel_1i2; hutel_20b19; hutel_20g21; hutel_20h13; hutel_20m11; hutel_20m24; hutel_21d15; hutel_22d2; hutel_22e12; hutel_22n2; hutel_22o2; hutel_23e13; hutel_23g11; hutel_24c19; hutel_24e11; hutel_24j6; hutel_2h3; their complements; and variants thereof.

22. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; ; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_41m15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11;

hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4; hfbr1_10e4; hfbr2_82g14; hfbr1_10g14; hfbr2_82i17; hfbr1_10; hfbr2_82i24; hfbr1_10; hfbr2_82m16; hfbr1_10; hfbr2_82m6; hfbr1_10; complements of the nucleic acid sequences; and variants thereof.

23. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16f21; hfbr2_16k22; hfbr2_22f21; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23f2; ; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2c1; hfbr2_2c18; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2k19; hfbr2_3f16; hfbr2_3l2; hfbr2_62n10; hfbr2_64a11; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6i20; hfbr2_71o20; hfbr2_72d13; hfbr2_72m16; hfbr2_72n12; hfbr2_78d13; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82m16; hfbr1_10; complements of the nucleic acid sequences; and variants thereof.

24. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; complements of the nucleic acid sequences; and variants thereof.

25. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfkd2_1j9; hfkd2_24e23; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_4b6; hfkd2_4c8; complements of the nucleic acid sequences; and variants thereof.

26. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of:

hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; complements of the nucleic acid sequences; and variants thereof.

27. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hmcfl_1c23; hmcfl_1g13; complements of the nucleic acid sequences; and variants thereof.

28. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hhtes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17l17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18l7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21l16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23l11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2l19; htes3_2m18; htes3_2m20; htes3_2n9; htes3_2o13; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06; htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; complements of the nucleic acid sequences; and variants thereof.

29. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: htes3_14g5; htes3_14p14; htes3_14p7; htes3_15a13; htes3_15g14; htes3_15h1; htes3_15j18; htes3_17f10; htes3_17n18; Htes3_18f3; htes3_19f19; htes3_19j17; htes3_20c21; htes3_21n23; htes3_22c23; htes3_22n13; Htes3_23n19; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2d15; htes3_2f14; htes3_2g7; htes3_2h15; htes3_2l19; htes3_2m20; htes3_2n9; htes3_30f4; htes3_35g6; htes3_35n24; htes3_35p17; htes3_4b4; htes3_4f17;

htes3_4o19; htes3_50j4; htes3_50n23; htes3_50n06; htes3_6b21; htes3_6d16; htes3_72k11; htes3_7d17; htes3_7j8; Htes3_8g11; Htes3_8g5; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; complements of the nucleic acid sequences; and variants thereof.

30. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16g18; hfbr2_2k14; Htes3_35b4; htes3_35p22; htes3_7j3; htes3_7p10; hute1_20m11; complements of the nucleic acid sequences; and variants thereof.

31. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16c16; hfbr2_2b5; htes3_15i5; htes3_18l7; htes3_1k11; Htes3_72k15; htes3_7b22; hute1_19g22; hute1_24j6; complements of the nucleic acid sequences; and variants thereof.

32. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_2d15; htes3_35e21; hute1_2h3; complements of the nucleic acid sequences; and variants thereof.

33. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_23l24; hfbr2_2i17; hfbr2_41m15; hfbr2_62f10; hfbr2_62l19; hfbr2_64j18; hfkd2_24n20; hfkd2_24p5; hfkd2_4k14; htes3_1g13; htes3_21l16; htes3_23l11; htes3_26g22; htes3_4h6; htes3_72p16; hute1_19h17; hute1_20h13; hute1_24e11; complements of the nucleic acid sequences; and variants thereof.

34. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_3g8; hfbr2_62o17; hfbr2_6b24; hfbr2_78k24; hfkd2_24b15; hfkd2_3o17; hfkd2_46j20; htes3_17l17; Htes3_17n18; htes3_27d1; htes3_2a17; htes3_35b5; htes3_35k16; htes3_35n12; htes3_35n9; hute1_20b19; hute1_20m24; hute1_23e13; complements of the nucleic acid sequences; and variants thereof.

35. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of:

hfbr2_23b10; hfbr2_3c18; hfbr2_64a15; hfbr2_6o17; hfbr2_72b18; hfbr2_72l12; hfbr2_82i24(hfbr1_10); hte3_14h21; Htes3_15j3; htes3_20m18; htes3_22g2; htes3_2m18; htes3_7p9; htes3_8m10; hutel_18l1; complements of the nucleic acid sequences; and variants thereof.

36. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_23b21; hfbr2_23n16; hfbr2_2c17; hfbr2_62b11; hfbr2_78c24; hfbr2_82e4 (hfbr1_10e4); hfbr2_82i17 (hfbr1_10); hfbr2_82m6 (hfbr1_10); hfkd2_46m4; htes3_15k11; htes3_1c1; hhtes3_1n3; htes3_20k2; htes3_21d4; htes3_23n19; htes3_4f5; htes3_6c11; htes3_8e24; hutel_20g21; hutel_22d2; hutel_22e12; complements of the nucleic acid sequences; and variants thereof.

37. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfbr2_16i12; hfbr2_16l12; hfbr2_22h13; hfbr2_2b17; hfbr2_2d17; hfbr2_64k24; hfbr2_82c20 (hfbr1_10c20); hfbr2_82e17 (hfbr1_10e17); hfbr2_82g14 (hfbr1_10g14); hfkd2_24a15; hfkd2_3i13; hfkd2_4m11; hmcf1_1a11; hmcf1_1e15; htes3_15c6; htes3_2o13; htes3_27k4; htes3_2h1; htes3_35k24; hutel_19f19; and hutel_24c19; complements of the nucleic acid sequences; and variants thereof.

38. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfkd2_46k19; hfkd2_47a4; htes3_2e12; htes3_21j15; htes3_17n12; hutel_18i19; hutel_1i2; complements of the nucleic acid sequences; and variants thereof.

39. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hutel_17k7; hutel_18c12; hutel_18i19; hutel_18i4; hutel_18l1; hutel_19f19; hutel_19g19; hutel_19g22; hutel_19h17; hutel_19j11; hutel_1i2; hutel_20b19; hutel_20g21; hutel_20h13; hutel_20m11; hutel_20m24; hutel_21d15; hutel_22d2; hutel_22e12; hutel_22n2; hutel_22o2; hutel_23e13; hutel_23g11; hutel_24c19; hutel_24e11; hutel_24j6; hutel_2h3; complements of the nucleic acid sequences; and variants thereof.

40. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hutel_17k7; hutel_18c12; hutel_18i4; hutel_19g19; hutel_19j11; hutel_22n2; hutel_21d15; hutel_22o2; hutel_23g11; complements of the nucleic acid sequences; and variants thereof.

41. A nucleic acid molecule having the sequence of a clone selected from the group consisting of hfbr2_16c16; hfbr2_16f21; hfbr2_16g18; hfbr2_16i12; hfbr2_16k22; hfbr2_16l12; hfbr2_22f21; hfbr2_22h13; hfbr2_22h13; hfbr2_22i4; hfbr2_22k3; hfbr2_22k8; hfbr2_23b10; hfbr2_23b21; hfbr2_23f2; hfbr2_23l24; hfbr2_23n16; hfbr2_23o24; hfbr2_23o5; hfbr2_2a2; hfbr2_2b17; hfbr2_2b5; hfbr2_2c1; hfbr2_2c17; hfbr2_2c18; hfbr2_2d15; hfbr2_2d17; hfbr2_2d20; hfbr2_2g18; hfbr2_2h1; hfbr2_2h10; hfbr2_2i17; hfbr2_2k14; hfbr2_2k19; hfbr2_3b16; hfbr2_3c18; hfbr2_3f16; hfbr2_3g8; hfbr2_3l2; hfbr2_41m15; hfbr2_62b11; hfbr2_62f10; hfbr2_62l19; hfbr2_62n10; hfbr2_62o17; hfbr2_64a11; hfbr2_64a15; hfbr2_64c16; hfbr2_64c4; hfbr2_64h6; hfbr2_64i20; hfbr2_64j18; hfbr2_64k24; hfbr2_64o16; hfbr2_6a17; hfbr2_6b24; hfbr2_6i20; hfbr2_6o17; hfbr2_71o20; hfbr2_72b18; hfbr2_72d13; hfbr2_72l12; hfbr2_72m16; hfbr2_72n12; hfbr2_78c24; hfbr2_78d13; hfbr2_78k24; hfbr2_78n23; hfbr2_7a24; hfbr2_7e22; hfbr2_7j4; hfbr2_82c20; hfbr1_10c20; hfbr2_82e17; hfbr1_10e17; hfbr2_82e4;; hfbr1_10e4; hfbr2_82g14;; hfbr1_10g14; hfbr2_82i17;; hfbr1_10; hfbr2_82i24;; hfbr1_10; hfbr2_82m16;; hfbr1_10; hfbr2_82m6;; hfbr1_10; hfkd2_1j9; hfkd2_24a15; hfkd2_24b15; hfkd2_24e23; hfkd2_24n20; hfkd2_24p5; hfkd2_3i13; hfkd2_3o17; hfkd2_46a6; hfkd2_46b10; hfkd2_46d13; hfkd2_46j20; hfkd2_46k19; hfkd2_46m4; hfkd2_47a4; hfkd2_4b6; hfkd2_4c8; hfkd2_4k14; hfkd2_4m11; hmcfl_1a11; hmcfl_1c23; hmcfl_1e15; hmcfl_1g13; hhtes3_1n3; htes3_14g5; htes3_14h21; htes3_14p14; htes3_14p7; htes3_15a13; Htes3_15c24; htes3_15c6; htes3_15g14; htes3_15h1; htes3_15i5; htes3_15j18; Htes3_15j3; htes3_15k11; htes3_17f10; htes3_17l17; htes3_17n12; htes3_17n18; Htes3_18f3; htes3_18l7; htes3_19f19; htes3_19j17; htes3_1c1; htes3_1g13; htes3_1k11; htes3_20c21; htes3_20k2; htes3_20m18; htes3_21d4; htes3_21j15; htes3_21l16; htes3_21n23; htes3_22c23; htes3_22g2; htes3_22n13; htes3_23l11; htes3_23n19; Htes3_23n19; htes3_26g22; htes3_27d1; htes3_27k4; htes3_27o14; htes3_28d14; htes3_2a11; htes3_2a17; htes3_2d15; htes3_2e12; htes3_2f14; htes3_2g7; htes3_2h1; htes3_2h15; htes3_2l19; htes3_2m18;

htes3_2m20; htes3_2n9; htes3_2ol3; htes3_30f4; Htes3_35b4; htes3_35b5; htes3_35e21; htes3_35g6; htes3_35k16; htes3_35k24; htes3_35n12; htes3_35n24; htes3_35n9; htes3_35p17; htes3_35p22; htes3_4b4; htes3_4f17; htes3_4f5; htes3_4h6; htes3_4o19; htes3_50j4; htes3_50n06; htes3_50n23; htes3_6b21; htes3_6c11; htes3_6d16; htes3_72k11; Htes3_72k15; htes3_72p16; htes3_7b22; htes3_7d17; htes3_7j3; htes3_7j8; htes3_7p10; htes3_7p9; htes3_8e24; Htes3_8g11; Htes3_8g5; htes3_8m10; Htes3_8p7; Htes3_9e22; Htes3_9i20; Htes3_9k22; hutel_17k7; hutel_18c12; hutel_18i19; hutel_18i4; hutel_18l1; hutel_19f19; hutel_19g19; hutel_19g22; hutel_19h17; hutel_19j11; hutel_1i2; hutel_20b19; hutel_20g21; hutel_20h13; hutel_20m11; hutel_20m24; hutel_21d15; hutel_22d2; hutel_22e12; hutel_22n2; hutel_22o2; hutel_23e13; hutel_23g11; hutel_24c19; hutel_24e11; hutel_24j6; hutel_2h3; their complements; and variants thereof.

42. A polypeptide encoded by the nucleic acid molecule according to claim 41.

43. An antibody or fragment thereof that is capable of binding to a specific portion of the peptide according to claim 42.

44. A pharmaceutical composition, comprising (a) an effective amount of a pharmaceutical agent, wherein said pharmaceutical agent is selected from the group consisting of the polypeptide according to claim 42, variants or functional derivatives thereof, and antibodies thereto; and (2) a physiologically acceptable carrier or excipient.

45. An expression vector comprising the nucleic acid molecule of claim 41 or a fragment thereof, and optionally a promoter operably linked to said nucleic acid molecule or said fragment.

46. A method for recombinantly producing a desired peptide, comprising expressing in a host cell a peptide encoded by the nucleic acid molecule according to claim 41.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12659 A3

(51) International Patent Classification[7]: C12N 15/12, C07K 14/47, C12Q 1/68, C07K 16/18, A61K 38/17, C12P 21/00

(21) International Application Number: PCT/IB00/01496

(22) International Filing Date: 18 August 2000 (18.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,499 18 August 1999 (18.08.1999) US
60/156,503 28 September 1999 (28.09.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/156,503 (CIP)
Filed on 18 August 1999 (18.08.1999)

(71) Applicant *(for all designated States except US)*: **FRAUNHOFER-GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE]; Leonrodstrasse 54, D-80636 München (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **WIEMANN, Stefan** [DE/DE]; Grosse Lachstrasse 30a, 69207 Sandhausen (DE). **POUSTKA, Annemarie** [DE/DE]; Ladenburgstrasse 41, 69120 Heidelberg (DE). **WELLENREUTHER, Ruth** [DE/DE]; An der Markscheide 5, 69126 Heidelberg (DE). **BLUM, Helmut** [DE/DE]; Koenigswieser Strasse 94, 81475 Muenchen (DE). **OBERMAIER, Brigitte** [DE/DE]; Muehlstrasse 9a, 82547 Eurasberg (DE). **OTTENWAELDER, Birgit** [DE/DE]; Beinhoferstrasse 1a, 81247 Muenchen (DE). **BAHR, André** [DE/DE]; Raffaelweg 6, 40724 Hilden (DE). **DUESTERHOEFT, Andreas** [DE/DE]; Karlrobert-Kreiten-Strasse 14, 40724 Hilden (DE). **KOENIG, Christoph** [DE/US]; 6233 22nd Avenue N.E., Seattle, WA 98115 (US). **LAUBER, Juergen** [DE/DE]; Unterberg 1F, 42799 Leichlingen (DE). **HEUBNER, Dagmar** [DE/DE]; Gruene Trift 126a, 12557 Berlin (DE). **WAMBUTT, Rolf** [DE/DE]; Florian-Geyer-Strasse 28, 12489 Berlin (DE). **KOEHRER, Karl** [DE/DE]; Schlossmannstrasse 4, 40225 Duesseldorf (DE). **BEYER, Andreas** [DE/DE]; Helgolandring 106, 45149 Essen (DE). **GASSENHUBER, Johann** [DE/DE]; Emanuel Geibel Strasse 8, 65185 Wiesbaden (DE). **GRUBER, Christian** [DE/DE]; Zasinger Strasse 8, 81547 Muenchen (DE). **STRACK, Norman** [DE/DE]; Linderbergweg 1, 82229 Seefeld (DE). **MEWES, H.W.** [DE/DE]; Graf Toerring Strasse 9, 82237 Woerthsee (DE). **ANSORGE, Wilhelm** [DE/DE]; Boxberring 107/55, 69126 Heidelberg (DE). **GLASSL, Sabine** [DE/DE]; Friedberger Weg 2, 64720 Michelstadt (DE). **RITTMUELLER, Claudia** [DE/DE]; Siedlerweg 2, 69151 Dilsbergerhof (DE). **REGIERT, Thomas** [DE/DE]; Raiffeisenstrasse 38, 67227 Frankenthal (DE). **BLOECKER, Helmut** [DE/DE]; Doeringstrasse 16, 38118 Braunschweig (DE). **BOECHER, Michael** [DE/DE]; Alter Weg 41a, 38302 Wolfenbuettel (DE). **HORNISCHER, Klaus** [DE/DE]; Mozartstrasse 2, 38106 Braunschweig (DE). **NORDSIEK, Gabriele** [DE/DE]; Ohefeld 34, 31188 Holle (DE). **TAMPE, Jens** [DE/DE]; Bergisch-Gladbacher-Strasse 656, 51067 Koeln (DE).

(74) Agents: **MERCER, Christopher, Paul** et al.; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 20 June 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/12659 A3

(54) Title: HUMAN DNA SEQUENCES

(57) Abstract: Novel human cDNA sequence of a clones, the encoded protein sequence of a clones, antibodies and variants thereof, are provided. The disclosed sequence of a clones find application in a number of ways, including use in profiling assays. In this regard, various assemblages of nucleic acids or proteins are provided that are useful in providing large arrays of human material for implementing large-scale screening strategies. The disclosed sequence of a clones may also be used in formulating medicaments, treating various disorders and in certain diagnostic applications.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 00/01496

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/47 C12Q1/68 C07K16/18 A61K38/17
C12P21/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 00 09552 A (GENETICS INST) 24 February 2000 (2000-02-24) Page 546, claim 86: SEQ.ID.No.: 77 ---	1-46
X	HILLIER L ET AL: "Human cDNA clone IMAGE:754267" EMBL SEQUENCE DATABASE, 23 July 1997 (1997-07-23), XP002163418 HEIDELBERG DE Accession Nr.: AA478899 abstract ---	1-42

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 March 2001	07.06.01
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer De Kok, A

2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 00/01496

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HILLIER L ET AL.: "Human cDNA clone IMAGE: 754167" EMBL SEQUENCE DATABASE, 23 June 1997 (1997-06-23), XP002163419 HEIDELBERG DE Accession Nr.: AA478780 abstract ---	1-42
X	STRAUSBERG R ET AL.: "Human cDNA sequence IMAGE:2138166" EMBL SEQUENCE DATABASE, 24 March 1999 (1999-03-24), XP002163420 HEIDELBERG DE Accession Nr.:522149 abstract ---	1-42
X	HILLIER L ET AL.: "Human cDNA clone IMAGE:263887" EMBL SEQUENCE DATABASE, 5 January 1996 (1996-01-05), XP002163421 HEIDELBERG DE Accession Nr.: N28525 abstract ---	1-42
A	"Atlas(tm) human cDNA expression array I" CLONTECHNIQUES,April 1977 (1977-04), pages 4-7, XP002914393 US the whole document ---	1-20
A	REICHERT J ET AL: "HUMAN AND RODENT EXPRESSION PATTERN OF A FUSION GENE ISOLATED FROM AN MCF7 CDNA LIBRARY" INTERNATIONAL JOURNAL OF ONCOLOGY, vol. 9, no. 1, 1996, pages 29-32, XP000906725 page 29 ---	1,6,7,17
A	WO 98 40486 A (GENETICS INST) 17 September 1998 (1998-09-17) page 29, line 20 -page 60, line 13 page 18, line 5 -page 26, line 32 -----	1-5, 8-25, 28-46

International application No.
PCT/IB 00/01496

INTERNATIONAL SEARCH REPORT

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 21-40
because they relate to subject matter not required to be searched by this Authority, namely:

Rule 39.1(v) PCT - Presentation of information:
Although claims 21-40 could be considered as a mere presentation of information, according to Rule 39.1(v) PCT, the search has been carried out as far as possible in our systematic documentation.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-46 all partially

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-46, all partially

Invention 1:

A nucleic acid molecule having the sequence of the clone hfbr2_16c16 (corresponding to SEQ.ID.1); an assemblage comprising said nucleic acid; a computer readable medium comprising said nucleic acid; a polypeptide encoded by said nucleic acid; an antibody binding to said polypeptide; an expression vector comprising said nucleic acid and a method for producing said polypeptide.

2. Claims: 1-46, all partially

Invention 2-233:

same as invention 1, but for each single clone as set forth in claim 1 (i.e. starting with clone hfbr2_16f21 and ending with clone hutel_2h3)

NB: for the sake of conciseness, the first subject-matter is explicitly defined, the other subject-matter by analogy thereto.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/IB 00/01496

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0009552 A	24-02-2000	AU 5557099 A	06-03-2000
WO 9840486 A	17-09-1998	US 5976837 A	02-11-1999
		AU 6702298 A	29-09-1998
		EP 0973890 A	26-01-2000